МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ФАКУЛЬТЕТ МАТЕМАТИКИ, ИНФОРМАЦИОННЫХ
И АВИАЦИОННЫХ ТЕХНОЛОГИЙ
КАФЕДРА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И ТЕОРИИ УПРАВЛЕНИЯ

Рацеев С.М.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ

Электронное учебное пособие

Ульяновск
2018

УДК 004.056:51(075.8)
ББК 32.972.53я04я73
Р 27

Рацеев С.М. Математические методы защиты информации [Электронный ресурс]: Электронное учеб. пособие/ С.М. Рацеев. Ульяновск: УлГУ, 2018. 1 CD-R. № гос. регистрации – 0321901084.

В данном учебном пособии изложены некоторые подходы и методы современной криптографии. Рассмотрены математические основы криптографии, математические модели шифров, конструкции невскрываемых криптосистем, симметричные блочные шифры, основные шифры с открытыми ключами, криптографические хеш-функции, коды аутентификации, методы электронной подписи, схемы разделения секрета, протоколы аутентификации, протоколы с нулевым разглашением знания, протоколы передачи ключей, криптосистемы на эллиптических кривых. Подробно рассмотрены криптосистемы, лежащие в основе криптографических отечественных стандартов.

Работа предназначена для студентов, изучающих дисциплины «Криптографические методы защиты информации», «Криптографические протоколы», «Методы алгебраической геометрии в криптографии», «Теоретико-числовые методы в криптографии» по специальностям «Компьютерная безопасность» и «Информационная безопасность автоматизированных систем».

УДК 004.056:51(075.8)
ББК 32.972.53я04я73

Оглавление

Введение **12**

1 Элементы теоретико-числовых методов в криптографии **19**

1.1 Отношение делимости целых чисел. Наибольший общий делитель 19
1.2 Обобщенный алгоритм Евклида 23
1.3 Наименьшее общее кратное 28
1.4 Диофантовы уравнения первой степени 30
1.5 Системы диофантовых уравнений первой степени 34
1.6 Простые числа 40
1.7 Разложение $n!$ на простые множители 46
1.8 Цепные дроби 49
1.8.1 Конечные цепные дроби 49
1.8.2 Бесконечные цепные дроби 58
1.8.3 Квадратичные иррациональности 66
1.9 Мультипликативные функции 69
1.9.1 Функция Мебиуса 71
1.9.2 Функция Эйлера 72
1.10 Сравнения 74
1.11 Теоремы Эйлера и Ферма 78
1.12 Сравнения первой степени 80
1.13 Решение линейных диофантовых уравнений с использованием сравнений 82
1.14 Системы сравнений первой степени 85
1.15 Сравнения по простому модулю 88
1.16 Сравнения по составному модулю 94
1.17 Степенные вычеты 96

1.18 Первообразные корни по простому модулю 99
1.19 Первообразные корни по составному модулю . . . 101
1.20 Индексы (дискретные логарифмы) 107
1.21 Сравнения второй степени 111
1.21.1 Квадратичные вычеты и невычеты 111
1.21.2 Символ Лежандра 115
1.21.3 Символ Якоби 122
1.22 Вычисление квадратного корня 125
1.23 Тесты на простоту 130
1.23.1 Тест на основе малой теоремы Ферма . . . 130
1.23.2 Тест Соловея-Штрассена 133
1.23.3 Тест Миллера-Рабина 137
1.23.4 $N-1$ методы доказательства простоты . . 143
1.24 Разложение целых чисел на множители 148
1.24.1 Метод пробного деления 148
1.24.2 ρ-метод Полларда 149
1.24.3 $(p-1)$-метод Полларда 150
1.25 Методы дискретного логарифмирования 152
1.25.1 Полный перебор 153
1.25.2 Метод Гельфонда-Шенкса 154
1.25.3 ρ-метод Полларда 155
1.25.4 Метод исчисления порядка 158
1.25.5 Решение систем линейных сравнений 161

2 Алгебраические основы криптографии 165
2.1 Бинарные отношения 165
2.1.1 Отношение эквивалентности 165
2.1.2 Отношение частичного порядка 167
2.2 Алгебраические операции 172
2.3 Полугруппы. Группы 174
2.3.1 Подгруппы 177
2.3.2 Смежные классы. Теорема Лагранжа 182
2.3.3 Нормальная подгруппа 185
2.3.4 Фактор-группа 185
2.3.5 Морфизмы групп 186

2.4 Кольца 190
2.4.1 Кольца многочленов 193
2.4.2 Подкольца 195
2.4.3 Идеалы кольца 197
2.4.4 Фактор-кольцо 197
2.4.5 Кольцо классов вычетов 198
2.4.6 Морфизмы колец 199
2.4.7 Кольца главных идеалов 200
2.4.8 Китайская теорема об остатках 201
2.5 Поля 204
2.5.1 Простые идеалы 205
2.5.2 Подполе. Поле частных 206
2.5.3 Простые поля. Характеристика поля 211
2.5.4 Расширение полей 212
2.5.5 Поля разложения многочлена 218
2.5.6 Конечные поля 222
2.5.7 Образующие элементы конечного поля . . . 225
2.5.8 Неприводимые многочлены над конечными полями 227
2.5.9 Автоморфизм Фробениуса. Совершенные поля 231
2.5.10 Трансцендентные расширения полей 232

3 Элементы алгебраической геометрии 237
3.1 Аффинные алгебраические многообразия 237
3.2 Проективная плоскость 245
3.3 Эллиптические кривые 254
3.4 Сложение точек эллиптической кривой над произвольным полем 265

4 Математические модели открытых текстов 268
4.1 Детерминированная модель 268
4.2 Вероятностная модель 270
4.2.1 Вероятностная модель независимых символов алфавита 271

4.2.2 Вероятностная модель независимых биграмм . . . 272
4.2.3 Вероятностная модель марковски зависимых букв . . . 273

5 Шифры замены и перестановки (исторические шифры) 275
5.1 Одноалфавитные шифры замены . . . 275
5.1.1 Шифр простой замены . . . 275
5.1.2 Шифр сдвига . . . 276
5.1.3 Улучшенный криптоанализ шифра сдвига . 277
5.1.4 Аффинный шифр . . . 280
5.1.5 Преобразование биграмм аффинным шифром . . . 281
5.2 Многоалфавитные шифры замены . . . 282
5.2.1 Шифр замены с конечным ключом . . . 282
5.2.2 Шифр Виженера . . . 283
5.2.3 Криптоанализ шифра Виженера . . . 284
5.2.4 Многопетлевые подстановки . . . 289
5.2.5 Аффинный блочный шифр . . . 289
5.2.6 Табличное гаммирование . . . 291
5.2.7 Модульное гаммирование . . . 292
5.2.8 Шифр пропорциональной замены . . . 293
5.3 Шифры перестановки . . . 296
5.3.1 Маршрутные перестановки . . . 297

6 Надежность шифров 299
6.1 Формальные модели шифров . . . 299
6.2 Математические модели некоторых шифров . . . 302
6.3 Ортогональные таблицы . . . 304
6.4 Совершенные шифры . . . 305
6.5 $(k|y)$-совершенные шифры . . . 317
6.6 Математические модели шифра замены с ограниченным и неограниченным ключом . . . 324
6.7 Совершенные шифры замены . . . 329
6.8 $(k|y)$-совершенные шифры Σ_H . . . 336

6.9 Вопросы имитостойкости шифров 338
6.10 Совершенные имитостойкие шифры 346
6.11 Совершенные имитостойкие шифры на основе ортогональных таблиц 355

7 Шифры, не распространяющие искажений 359
7.1 Шифры, не распространяющие искажений типа замены знаков 359
7.2 Шифры, не распространяющие искажений типа пропуска знаков 366
7.3 Шифры, не распространяющие искажений типа вставки знаков 374

8 Симметричные блочные шифры 376
8.1 Общие сведения 376
8.2 Итеративные блочные шифры 377
8.3 Шифры Фейстеля 379
8.4 Построение раундовой функции 381
8.5 Входное и выходное отображения 383
8.6 Слабые ключи итеративного блочного шифра . . . 384
8.7 Режимы использования блочных шифров 386
8.8 Стандарт симметричного блочного шифрования ГОСТ Р 34.12-2015 389
8.9 Шифр AES (Rijndael) 396
8.10 Основы криптоанализа 405
8.10.1 Криптоатаки 405
8.10.2 Метод полного перебора 406
8.10.3 Аналитический метод 410
8.10.4 Метод встречи посередине 411

9 Шифрование с открытым ключом 413
9.1 Предыстория и основные идеи 413
9.2 Система Диффи-Хеллмана и ее модификация на эллиптической кривой 417
9.3 Протокол Месси-Омуры и его модификация на эллиптической кривой 420

9.4 Вероятностный шифр Эль-Гамаля и его модификация на эллиптической кривой 423
9.5 Шифр RSA 426
9.6 Рюкзачная криптосистема Меркла-Хеллмана . . . 429
9.7 Рюкзачная криптосистема Шора-Ривеста на основе конечных полей 431

10 Криптографические хеш-функции 439
10.1 Сбалансированные функции 439
10.2 Хеш-функции и целостность данных 442
10.3 Криптографические хеш-функции 445
10.4 Построение хеш-функций 448
10.5 Хеш-функция ГОСТ Р 34.11-2012 450
10.6 Парадокс дней рождений 453

11 Коды аутентификации 463
11.1 Основные понятия 463
11.2 Оптимальные коды аутентификации 468
11.3 Математическая модель кода аутентификации с неограниченным ключом 473

12 Электронная подпись 479
12.1 Общие понятия 479
12.2 Электронная подпись RSA 481
12.3 Электронная подпись Фиата-Шамира 483
12.4 Электронная подпись Эль-Гамаля 484
12.5 Электронная подпись Шнорра 485
12.6 Электронная подпись на основе эллиптических кривых 486
12.6.1 Электронная подпись ГОСТ Р 34.10-2012 . 486
12.6.2 Электронная подпись ECDSA 490
12.7 Электронная подпись Диффи-Лампорта на основе симметричных систем шифрования 492

13 Схемы разделения секрета 494
13.1 Пороговые схемы разделения секрета 494

13.1.1 Схема разделения секрета Шамира 495
13.1.2 Проверяемая схема Фельдмана-Шамира и ее модификация на эллиптических кривых 498
13.1.3 Совершенная проверяемая схема Педерсена-Шамира и ее модификация на эллиптических кривых 500
13.1.4 Схема разделения секрета на основе СЛАУ 502
13.1.5 Схема разделения секрета на основе равновесных двоичных кодов 503
13.1.6 Схема разделения секрета на основе китайской теоремы об остатках 505
13.2 Схемы разделения секрета для произвольных структур доступа 507
13.2.1 Схема Бенало-Лейхтера 507
13.2.2 Схема Ито-Саито-Нишизеки 508

14 Протоколы аутентификации 510
14.1 Протоколы аутентификации, использующие пароли (слабая аутентификация) 511
14.2 Протоколы аутентификации, использующие технику «запрос–ответ» (сильная аутентификация) 513
14.2.1 «Запрос–ответ» с использованием симметричных алгоритмов шифрования . . 513
14.2.2 «Запрос–ответ» с использованием асимметричных алгоритмов шифрования . 515
14.3 Протоколы аутентификации с нулевым разглашением знания 517
14.3.1 Протокол аутентификации Фиата-Шамира 520
14.3.2 Протокол Фейга-Фиата-Шамира 522
14.3.3 Протокол аутентификации с нулевым разглашением без доверенного центра . . . 523
14.3.4 Протокол аутентификации Шнорра 525
14.3.5 Трехпроходный протокол аутентификации Шнорра 528
14.3.6 Протокол аутентификации Окамото 529

14.3.7 Модификация (усиление) протокола Шнорра 530
14.3.8 Модификация протоколов Шнорра и Окамото на эллиптических кривых 531
14.3.9 Протокол аутентификации Гиллоу-Куискатр (GQ) 534
14.3.10 Протокол аутентификации на основе задачи о доказательстве изоморфизма графов 535
14.3.11 Протокол аутентификации на основе задачи о раскраске графа 536
14.3.12 Протокол аутентификации на основе задачи о нахождении гамильтонова цикла в графе 539
14.3.13 Протоколы аутентификации на основе асимметричных шифров 542

15 Протоколы с нулевым разглашением 548
15.1 Протоколы привязки к биту 548
15.2 Применение протоколов привязки к биту в протоколах аутентификации с нулевым разглашением 550
15.3 Протоколы типа «подбрасывание монеты по телефону» 552
15.4 Аргумент с нулевым разглашением 553
15.5 Протоколы электронного голосования 556
15.5.1 Протокол голосования на основе протокола Шаума-Педерсена 556
15.5.2 Протокол Крамера-Франклина-Шонмейкерса-Янга и его модификация на эллиптических кривых 560

16 Протоколы передачи ключей 565
16.1 Передача ключей с использованием симметричного шифрования 565
16.1.1 Двусторонние протоколы 565

16.1.2 Трехсторонние протоколы 568
16.2 Передача ключей с использованием асимметричного шифрования 574
16.3 Открытое распределение ключей 576
16.4 Предварительное распределение ключей 584

Литература **587**

Введение

Криптография — область научных, прикладных, инженерно-технических исследований и практической деятельности, которая связана с разработкой криптографических средств защиты информации от угроз со стороны противника и/или нарушителя, а также анализом и обоснованием их криптографической стойкости. В настоящее время основными задачами криптографии являются обеспечение конфиденциальности, целостности, аутентификации, невозможности отказа, неотслеживаемости. В отличие от организационных и других способов защиты информации, под криптографическими понимаются такие, которые используют математические методы преобразования защищаемой информации. Криптография, с некоторой долей условности, делится на: криптосинтез и криптоанализ; криптография включает криптологию.

Защита информации опирается на совокупность методов и алгоритмов, которые позволяют защитить информацию от атак противника. Существенной частью этой совокупности являются криптоалгоритмы. Криптоалгоритмы — это такие алгоритмы преобразования информации, которые используют секретный ключ. При этом так называемый принцип Керкгоффса гласит, что надежность криптоалгоритма должна зависеть только от секретности ключа и не зависеть от секретности самого криптоалгоритма. Поэтому основной параметр качества криптоалгоритма — устойчивость к попыткам противника заполучить секретный ключ. Такая устойчивость в криптографии называется стойкостью. Для обоснования криптографической стойкости требуется математическая формализация задачи и математические модели исследуемых объектов. Имеются различные подхо-

ды к построению таких моделей. В данном пособии используются модели из работ [2, 22].

В данной работе используются криптографические термины из работы [32].

В структурном плане учебное пособие состоит из шестнадцати глав.

В главе 1 приводятся элементы теории чисел, лежащие в основе криптографии с открытым ключом. Рассматриваются следующие темы: отношение делимости в кольце целых чисел и его свойства; наибольший общий делитель и его свойства; алгоритм Евклида, обобщенный алгоритм Евклида; диофантовы уравнения первой степени; системы диофантовых уравнений первой степени; наименьшее общее кратное и его свойства; взаимно простые числа и их свойства; простые числа и их свойства; основная теорема арифметики; конечные цепные дроби, представление рационального числа конечной цепной дробью; бесконечные цепные дроби; подходящие дроби, их вычисление и основные свойства; мультипликативные функции и их свойства; функция Мебиуса и ее свойства; функция Эйлера и ее вычисление; отношение сравнимости в кольце целых чисел и его свойства; полная и приведенная системы вычетов и их свойства; теорема Эйлера, теорема Ферма; сравнения первой степени; системы сравнений первой степени и методы их решения, китайская теорема об остатках; сравнения произвольной степени по простому модулю; сравнения по составному модулю; степенные вычеты, показатель числа; первообразные корни по простому модулю; первообразные корни по составному модулю; сравнения второй степени, символ Лежандра, символ Якоби; вычисление квадратного корня по простому модулю; некоторые детерминированные и вероятностные тесты на простоту; некоторые алгоритмы факторизации; некоторые алгоритмы дискретного логарифмирования. В данной главе используются результаты работ [1, 7, 8, 10, 13, 14, 26, 28, 29].

В главе 2 рассматриваются алгебраические основы криптографии. Приводятся следующие темы: группы, основные

свойства группы; подгруппы, эквивалентные условия подгруппы; циклическая группа, свойства циклических групп; смежные классы, индекс подгруппы, теорема Лагранжа; нормальная подгруппа, эквивалентные условия нормальной подгруппы, фактор-группа; морфизмы групп, ядро и образ гомоморфизма, теорема о гомоморфизме групп; кольца, мультипликативная группа кольца, подкольца, критерий подкольца; идеал кольца, фактор-кольцо, кольца вычетов; морфизмы колец, ядро и образ гомоморфизма, теорема о гомоморфизме колец, кольца главных идеалов; китайская теорема об остатках для идеалов колец, разложение кольца вычетов в прямую сумму примарных колец; поле: определение и основные свойства, подполе, критерий подполя, критерий конечного подполя; простые и максимальные идеалы; поле частных; простые поля, характеристика поля; расширение поля, теорема о башне полей; алгебраические и трансцендентные элементы поля, простые расширения полей, теорема о классификации простых расширений полей; поле разложения многочлена; конечные поля, построение конечного поля; образующие элементы конечного поля; неприводимые многочлены над конечными полями; автоморфизм Фробениуса, совершенные поля; трансцендентные расширения полей. Данная глава основана на работах [13, 25].

В главе 3 рассматриваются элементы алгебраической геометрии. Приводятся следующие темы: аффинные алгебраические многообразия, неприводимые аффинные многообразия, проективная плоскость, эллиптические кривые: определение, общая форма Вейерштрасса эллиптической кривой, сложение точек эллиптической кривой над произвольным полем. В данной главе используются результаты работ [5, 6, 23, 24, 31, 34, 39, 40, 41, 47, 59].

В главе 4 вводятся математические модели открытых текстов. Рассматривается детерминированная модель, вероятностная модель независимых символов алфавита, вероятностная модель независимых биграмм, вероятностная модель марковски зависимых букв.

В главе 5 рассматриваются шифры замены и перестановки. Хорошо известно, что в криптографии существуют два основных типа преобразований — замены и перестановки, а все остальные являются комбинацией этих двух типов. В данной главе приводятся некоторые одноалфавитные и многоалфавитные (исторические) шифры замены и их криптоанализ.

В главе 6 исследуются совершенные, $(k|y)$-совершенные и имитостойкие шифры. К. Шеннон в 40-х годах XX века ввел понятие совершенного шифра, обеспечивающего наилучшую защиту открытых текстов. Такой шифр не дает криптоаналитику никакой дополнительной информации об открытом тексте на основе перехваченной криптограммы. При этом хорошо известный шифр гаммирования с равновероятной гаммой является совершенным, но максимально уязвимым к попыткам имитации и подмены. Это происходит потому, что в шифре гаммирования алфавиты для записи открытых и шифрованных текстов равномощны. Также в данном шифре должны использоваться равновероятные гаммы, что не всегда достигается на практике. В данной главе рассматриваются задачи построения совершенных и $(k|y)$-совершенных шифров по заданному набору параметров, приводятся необходимые и достаточные условия данных шифров, рассматриваются совершенные и $(k|y)$-совершенные шифры замены с неограниченным ключом, а также совершенные шифры, стойкие к имитации и подмене шифрованных сообщений с необязательно равномерным распределением на множестве ключей. Данная глава основана на работах [2, 22, 35, 36, 37].

В главе 7 исследуются шифры, не распространяющие искажений. Рассматриваются шифры, не распространяющие искажений типа замены знаков; шифры, не распространяющие искажений типа пропуска знаков; шифры, не распространяющие искажений типа вставки знаков. Приводятся необходимые и достаточные условия данных шифров. Все критерии приводятся с полными доказательствами. В качестве основы для данной главы стала работа [3].

В главе 8 рассматриваются вычислительно стойкие шифры с

секретным ключом. Такие шифры, в отличие от теоретически стойких (совершенных) шифров, могут быть вскрыты, но для этого требуется очень большое количество вычислений. Данные шифры обеспечивают шифрование и расшифрование данных с многократным превосходством в скорости по отношению к шифрам с открытыми ключами и имеют фиксированную длину ключа в отличие от теоретически стойких шифров, что и объясняет их широкое практическое применение. Приводятся блочные шифры «Магма» и «Кузнечик» из ГОСТ Р 34.12-2015, шифр AES. Данная глава основана на работах [4, 41, 43, 44].

В главе 9 рассматриваются асимметричные шифры. В симметричной криптографии каждая из переписывающихся сторон должна иметь копию общего секретного ключа, что создает сложнейшую проблему управления ключами. В криптосистемах, о которых пойдет речь в этой главе, используются два ключа: открытый и секретный. Открытый ключ может быть опубликован в общедоступном справочнике наряду с именем пользователя. В результате любой желающий может зашифровать с его помощью свое сообщение и послать закрытую информацию владельцу соответствующего секретного ключа. Расшифровать посланное сообщение сможет только тот, у кого есть секретный ключ. Рассматриваются следующие криптосистемы: система Диффи-Хеллмана, шифр Шамира, шифр Эль-Гамаля, шифр RSA, рюкзачная криптосистема Шора-Ривеста на основе конечных полей. В данной главе используются результаты работ [2, 4, 20, 41, 42].

В главе 10 исследуются криптографические хеш-функции. Приводятся требования, предъявляемые к хеш-функциям. Рассматриваются криптографические хеш-функции, способы построения криптографических хеш-функций.

В главе 11 рассматриваются коды аутентификации (без сокрытия). Особое внимание уделено оптимальным кодам аутентификации. Также приводится математическая модель кода аутентификации с неограниченным ключом. В данной главе используются работы [37, 38, 46].

В главе 12 рассматривается такое важное понятие в криптографии, как электронная подпись. Приводятся алгоритмы электронной подписи RSA, Фиата-Шамира, Эль-Гамаля, Шнорра, ГОСТ Р 34.10-2012, ECDSA. В данной главе используются результаты работ [2, 24, 41, 42].

В главе 13 приводятся следующие пороговые схемы разделения секрета: схема разделения секрета Шамира, схема разделения секрета на основе СЛАУ, схема разделения секрета на основе равновесных двоичных кодов, схема разделения секрета на основе китайской теоремы об остатках. В данной главе используются результаты работ [20, 46, 63, 64].

В главе 14 рассматриваются протоколы аутентификации. Особое внимание уделено протоколам аутентификации с нулевым разглашением знания: протокол Фиата-Шамира, протокол Фейга-Фиата-Шамира, итеративный протокол аутентификации Фиата-Шамира без доверенного центра, трехпроходный протокол аутентификации Фиата-Шамира без доверенного центра, итеративный протокол аутентификации Шнорра, трехпроходный протокол аутентификации Шнорра, протокол аутентификации Окамото, протокол аутентификации Гиллоу-Куискатр (GQ), протокол аутентификации с нулевым разглашением на основе доказательства изоморфизма графов, протокол аутентификации с нулевым разглашением на основе задачи о раскраске графа, протокол аутентификации с нулевым разглашением на основе задачи о нахождении гамильтонова цикла в графе, протокол аутентификации с нулевым разглашением на основе асимметричных шифров.

В главе 15 рассматриваются протоколы с нулевым разглашением: протокол подбрасывания монеты по телефону, протоколы привязки к биту, протоколы электронного голосования.

В главе 16 рассматриваются протоколы передачи ключей. Приводятся такие темы, как передача ключей с использованием симметричного шифрования, передача ключей с использованием асимметричного шифрования, открытое распределение ключей, предварительное распределение ключей.

В главах 14, 15, 16, среди прочих, используются результаты работ [9, 21, 27, 46].

Глава 1. Элементы теоретико-числовых методов в криптографии

1.1. Отношение делимости целых чисел. Наибольший общий делитель

Теорема 1.1 (**деление с остатком**). Для любых целых a и b, $b \neq 0$, существуют, и притом единственные, целые q и r, такие что $a = bq + r$, $0 \leqslant r < |b|$.

Доказательство. 1. Рассмотрим сначала случай $b > 0$. Рассмотрим последовательность:

$$b \cdot 1,\ b \cdot 2,\ b \cdot 3, \ldots \tag{1.1}$$

Если $a \geqslant 0$, то рассмотрим множество M таких чисел из последовательности (1.1), которые больше числа a. Поскольку всякое непустое подмножество натуральных чисел содержит наименьший элемент, обозначим через $b\widetilde{s}$ наименьший элемент множества M. Пусть $s = \widetilde{s} - 1$. Тогда $bs \leqslant a < b(s+1)$.

Если же $a < 0$, то $-a > 0$, и в качестве M возьмем множество таких чисел из последовательности (1.1), которые больше или равны значению $-a$. Обозначим через $b \cdot t$ наименьший элемент множества M. Тогда $b(t-1) < -a \leqslant bt$. Следовательно:

$$b(-t) \leqslant a < b(-t+1).$$

Таким образом, во всех случаях (относительно a) существует такое целое q, что:

$$bq \leqslant a < b(q+1).$$

Обозначим через r разность $a - bq$. Из (1.1) получаем, что:

$$0 \ \leqslant \ r = a - bq \ < \ b(q+1) - bq \ = \ b.$$

Поэтому $a = bq + r$ и $0 \leqslant r < b$.

2. Пусть теперь $b < 0$, a — произвольное. Из рассмотренных выше случаев следует, что найдутся такие целые q и r, для которых:

$$a = (-b)q + r, \quad 0 \leqslant r < -b = |b|.$$

Отсюда имеем:

$$a = b(-q) + r, \quad 0 \leqslant r < |b|.$$

Докажем теперь единственность таких q и r. Предположим, что найдутся еще такие $\widetilde{q}$ и $\widetilde{r}$, что:

$$a = bq + r, \quad 0 \leqslant r < |b|, \quad a = b\widetilde{q} + \widetilde{r}, \quad 0 \leqslant \widetilde{r} < |b|.$$

Тогда $bq + r = b\widetilde{q} + \widetilde{r}$. Пусть для определенности $\widetilde{r} > r$. Перепишем последнее равенство в такой форме: $\widetilde{r} - r = b(q - \widetilde{q})$. Так как $0 \leqslant r < \widetilde{r} < |b|$, то $0 < \widetilde{r} - r < |b|$. Поэтому $q \neq \widetilde{q}$, из чего следует, что $q \geqslant \widetilde{q} + 1$, если $b > 0$; $\widetilde{q} \geqslant q + 1$, если $b < 0$, т.е.:

$$\widetilde{r} - r = b(q - \widetilde{q}) \geqslant |b|.$$

Противоречие с тем, что $\widetilde{r} - r < |b|$. Поэтому $\widetilde{r} = r$, откуда следует, что $b(q - \widetilde{q}) = 0$. Так как $b \neq 0$, то $q - \widetilde{q} = 0$, т.е. $q = \widetilde{q}$. □

В теореме 1.1 число a называют *делимым*, b — *делителем*, q — *частным*, r — *остатком от деления*.

Теорема 1.2. Для любого целого $a > 0$ и целого $b \geqslant 2$ существует, и притом единственное, разложение вида:

$$\begin{gathered} a = a_n b^n + \ldots + a_1 b + a_0, \\ 0 \leqslant a_i < b, \ i = 0, \ldots, n-1, \ 0 < a_n < b. \end{gathered} \tag{1.2}$$

Доказательство. Применим индукцию по a. При $a = 1$ разложение (1.2) верно при $n = 0$ и $a_0 = 1$. Предположим, что разложение (1.2) верно для любого $s < a$, где $a \geqslant 2$. Из теоремы 1.1 следует, что число a можно представить в виде $a = qb + a_0$, где $0 \leqslant a_0 < b$. При этом в силу того, что $b \geqslant 2$, следует двойное неравенство $0 \leqslant q < a$. Если $q = 0$, то $0 < a_0 < b$, и равенство

(1.2) показано. Если же $q > 0$, то к q применим предположение индукции и тем самым получим равенство (1.2).

Покажем единственность разложения вида (1.2). Предположим, что имеется еще такое разложение:

$$a = x_m b^m + \ldots + x_1 b + x_0,$$

$$0 \leqslant x_i < b, \quad i = 0, \ldots, m-1, \quad 0 < x_m < b.$$

Тогда:

$$(a_n b^{n-1} + \ldots + a_1)b + a_0 = (x_m b^{m-1} + \ldots + x_1)b + x_0.$$

Из теоремы 1.1 следует, что $a_0 = x_0$ и:

$$a_n b^{n-1} + \ldots + a_1 = x_m b^{m-1} + \ldots + x_1.$$

Применяя к последнему равенству предположение индукции, получаем однозначность разложения (1.2). □

Представление числа a в виде (1.2) называется *представлением числа в b-ичной системе счисления* и записывается $a = (a_n \ldots a_0)_b$.

Определение 1.1. Пусть a и b — некоторые целые числа, $b \neq 0$. Число b называется делителем числа a, если существует такое целое число q, что выполняется равенство $a = bq$. При этом a называется кратным числа b, а q — частным от деления a на b. Делитель называется собственным, если он отличен от самого числа.

Если число b является делителем числа a, то для краткости будем писать $b|a$ (при этом подразумевается, что $b \neq 0$). Если же b не является делителем числа a, то будем писать $b \nmid a$.

Из определения 1.1 и теоремы 1.1 непосредственно следует такое утверждение.

Предложение 1.1. Пусть a и b — некоторые целые числа, причем $b \neq 0$. Число b является делителем числа a тогда и только тогда, когда остаток от деления a на b равен нулю.

В следующей теореме приведены основные свойства делимости.

Теорема 1.3 (свойства отношения делимости).

1. Для любого целого $a \neq 0$ справедливо $a|a$ (рефлексивность отношения делимости).

2. Для любого целого a справедливо $1|a$.

3. Если $b|a$, то при любом сочетании знаков $\pm b| \pm a$.

4. Если $c|b$ и $b|a$, то $c|a$ (транзитивность отношения делимости).

5. Если $b|a$, то для любого целого $k \neq 0$ справедливо $kb|ka$.

6. Если $kb|ka$, причем $k \neq 0$, тогда $b|a$.

7. Если $b|a$, то для любого целого c справедливо $b|ac$.

8. Если $c|a$ и $c|b$, тогда $c|(a+b)$ и $c|(a-b)$.

9. Если $c|a_1$, $c|a_2$,. . . , $c|a_n$ и b_1, b_2,. . . ,b_n — произвольные целые числа, тогда $c|(a_1b_1 + a_2b_2 + \ldots + a_nb_n)$.

10. Если $b_1|a_1$, $b_2|a_2$,. . . ,$b_n|a_n$, тогда $b_1b_2 \ldots b_n|a_1a_2 \ldots a_n$.

11. Если $b|a$ и $a \neq 0$, то $|a| \geqslant |b|$.

12. Если $b|a$ и $a|b$, то $|a| = |b|$.

Данные свойства являются простыми, поэтому **доказательство** теоремы предоставляется читателю.

Наибольший общий делитель. В дальнейшем будем рассматривать лишь положительные делители чисел. Всякое целое число, делящее одновременно целые числа a_1, a_2,. . . ,a_n, называется *общим делителем*. Наибольший из общих делителей называется *наибольшим общим делителем* и обозначается символом $(a_1, a_2, \ldots, a_n)$. Если $(a_1, a_2, \ldots, a_n) = 1$, то числа a_1, a_2,. . . ,a_n называются *взаимно простыми*. Числа a_1, a_2,. . . ,a_n называются *попарно взаимно простыми*, если $(a_i, a_j) = 1$ при $i \neq j$.

Ясно, что если числа a_1, a_2,. . . ,a_n попарно взаимно просты, то они взаимно просты.

Предложение 1.2. Если $b|a$ и $b > 0$, тогда совокупность общих делителей чисел a и b совпадает с совокупностью делителей числа b, в частности, $(a, b) = b$.

Доказательство. Очевидно, что всякий общий делитель чисел a и b является делителем и одного b. Обратно, пусть d — некоторый делитель числа b. Так как $b|a$, то по теореме 1.3

(пункт 4) следует, что d является и делителем числа a, поэтому d будет являться общим делителем чисел a и b. Таким образом, совокупность общих делителей чисел a и b совпадает с совокупностью делителей числа b. Поскольку наибольшим делителем числа b является само b, то $(a, b) = b$. □

Предложение 1.3. Если $a = bq + r$, тогда совокупность общих делителей чисел a и b совпадает с совокупностью общих делителей чисел b и r, в частности, $(a, b) = (b, r)$.

Доказательство. Пусть d — некоторый общий делитель чисел a и b. Так как $r = a - bq$, то d является делителем и числа r (теорема 1.3), поэтому d является общим делителем чисел b и r. Обратно, равенство $a = bq + r$ показывает, что любой общий делитель чисел b и r является делителем и числа a, поэтому является общим делителем чисел a и b. Следовательно, совокупность общих делителей чисел a и b совпадает с совокупностью общих делителей чисел b и r. В частности, должны совпадать и наибольшие элементы данных совокупностей, т.е. $(a, b) = (b, r)$. □

1.2. Обобщенный алгоритм Евклида

Для разыскания наибольшего общего делителя применяется следующий алгоритм, носящий название *алгоритма Евклида*. Пусть a и b — некоторые целые числа, $b \neq 0$. Согласно теореме 1.1 имеет место такая цепочка равенств:

$$\begin{array}{rll} a = & r_{-1} = bq_0 + r_1, & 0 < r_1 < |b|, \\ b = & r_0 = r_1 q_1 + r_2, & 0 < r_2 < r_1, \\ & r_1 = r_2 q_2 + r_3, & 0 < r_3 < r_2, \\ & \dots & \\ & r_{n-2} = r_{n-1} q_{n-1} + r_n, & 0 < r_n < r_{n-1}, \\ & r_{n-1} = r_n q_n, & \end{array} \tag{1.3}$$

которая заканчивается, когда некоторое r_{n+1} будет равным нулю. Это обязательно должно произойти, поскольку цепочка

неравенств:

$$|b| > r_1 > r_2 > \ldots > r_n > 0$$

не может содержать более $|b|$ положительных чисел.

Из предложений 1.2 и 1.3 следует, что совокупность общих делителей чисел a и b совпадает с совокупностью общих делителей чисел b и r_1,..., совпадает с совокупностью общих делителей чисел r_{n-1} и r_n, наконец, совпадает с совокупностью делителей числа r_n. В частности:

$$(a, b) = (b, r_1) = (r_1, r_2) = \ldots = (r_{n-1}, r_n) = r_n.$$

Из данного результата следует теорема 1.4.

Теорема 1.4. Совокупность общих делителей чисел a и b совпадает с совокупностью делителей их наибольшего общего делителя.

Теорема 1.5. Для любого набора целых чисел a_1, a_2,..., a_n верны следующие утверждения:

(i) $(a_1, a_2, \ldots, a_n) = ((\ldots((a_1, a_2), a_3), \ldots), a_n)$;

(ii) совокупность общих делителей чисел a_1, a_2,..., a_n совпадает с совокупностью делителей числа $(a_1, a_2, \ldots, a_n)$.

Доказательство проведем с помощью индукции по n. База индукции при $n = 2$ следует из теоремы 1.4.

Предположим, что теорема верна для всех $k < n$, где $n \geqslant 3$. Из предположения индукции следует, что совокупность общих делителей чисел a_1, a_2,..., a_n совпадает с совокупностью общих делителей чисел $(a_1, \ldots, a_{n-1})$ и a_n, из чего, в свою очередь, следует (теорема 1.4), что данная совокупность совпадает с делителями числа $((a_1, \ldots, a_{n-1}), a_n)$. Поэтому:

$$(a_1, a_2, \ldots, a_n) = ((a_1, \ldots, a_{n-1}), a_n). \qquad \square$$

Теорема 1.6 (Ламе). Количество операций деления, необходимых для вычисления наибольшего общего делителя двух натуральных чисел с помощью алгоритма Евклида, не превышает пятикратного количества цифр в десятичной записи меньшего из этих двух чисел.

Доказательство. Пусть $b \leqslant a$ — натуральные числа и b записывается m цифрами в десятичной системе счисления, т.е. $10^{m-1} \leqslant b < 10^m$. Докажем, что алгоритм Евклида, примененный к числам a и b, выполнит не более $5m$ делений с остатком.

Пусть $r_{-1} = a$ и $r_0 = b, r_1, \ldots, r_n$ — последовательность делителей в алгоритме Евклида. Тогда:

$$r_{k-1} = r_k q_k + r_{k+1},\ 0 < r_{k+1} < r_k,\ k = 0, 1, \ldots, n-1,\ r_n = (a, b).$$

Покажем, что:

$$r_{n-k} \geqslant \lambda^k, \quad k = 0, 1, \ldots, n, \tag{1.4}$$

где $\lambda = \dfrac{1+\sqrt{5}}{2} \approx 1.61\ldots$ есть корень квадратного уравнения $x^2 - x - 1 = 0$. Применим метод математической индукции. При $k = 0$ получаем $r_n \geqslant 1 = \lambda^0$. При $k = 1$ имеем:

$$r_{n-1} \geqslant r_n + 1 \geqslant 2 > \lambda.$$

Предположим, что неравенство (1.4) выполняется для любого $s < k$ при $s \geqslant 1$. Тогда:

$$r_{n-k} = r_{n-(k-1)} q_{n-(k-1)} + r_{n-(k-2)} \geqslant r_{n-(k-1)} + r_{n-(k-2)} \geqslant$$
$$\geqslant \lambda^{k-1} + \lambda^{k-2} = \lambda^k.$$

В последнем равенстве учитывалось равенство $\lambda^2 = 1 + \lambda$. Неравенство (1.4) доказано.

Так как $10^m > b$ и $b = r_0 \geqslant \lambda^n$, то:

$$m > n \lg \lambda > n/5.$$

В последнем неравенстве использована оценка:

$$\lambda > 10^{1/5} \approx 1.58\ldots.$$

Осталось заметить, что количество делений t в алгоритме (1.3) равно $n + 1$. Поэтому $t - 1 < 5m$ и $t < 5m + 1$. □

Замечание 1.1. Оценка теоремы Ламе достижима. Например, $a = 13$, $b = 8$.

Теорема 1.7. Пусть $r_1, \ldots, r_n$, $q_0, \ldots, q_n$ — последовательности остатков и неполных частных в алгоритме Евклида (1.3) для чисел a и b. Тогда для любого $k = 1, \ldots, n$ выполнено равенство:

$$r_k = ax_k + by_k,$$

где x_k и y_k — целые числа, определенные рекуррентными соотношениями:

$$x_k = x_{k-2} - x_{k-1}q_{k-1}, \quad y_k = y_{k-2} - y_{k-1}q_{k-1},$$

с начальными условиями:

$$x_0 = 0,\ x_1 = 1, \quad y_0 = 1,\ y_1 = -q_0.$$

Доказательство. При $k = 1$ из (1.3) видно, что утверждение верно. Предположим, что утверждение верно для любого $t < k$, $k \geqslant 2$. Тогда:

$$\begin{aligned} ax_k + by_k &= a(x_{k-2} - x_{k-1}q_{k-1}) + b(y_{k-2} - y_{k-1}q_{k-1}) = \\ &= ax_{k-2} + by_{k-2} - (ax_{k-1} + by_{k-1})q_{k-1} = \\ &= r_{k-2} - r_{k-1}q_{k-1}. \end{aligned}$$

Из (1.3) видно, что $r_{k-2} - r_{k-1}q_{k-1} = r_k$. □

Следствие 1.1. Для любых целых чисел a и b найдутся такие целые числа x и y, что:

$$ax + by = (a, b).$$

Вычисление чисел x_k, y_k удобно производить с помощью следующей таблицы:

k	-1	0	1	2	$\ldots$	n
q_k	$-$	q_0	q_1	q_2	$\ldots$	q_n
x_k	1	0	$x_{-1} - x_0q_0$	$x_0 - x_1q_1$	$\ldots$	$x_{n-2} - x_{n-1}q_{n-1}$
y_k	0	1	$y_{-1} - y_0q_0$	$y_0 - y_1q_1$	$\ldots$	$y_{n-2} - y_{n-1}q_{n-1}$

Пример 1.1. Пусть $a = 50$, $b = 27$. Найдем целые числа x и y, удовлетворяющие условию $ax + by = (a, b)$.

Сначала применим алгоритм Евклида:

$$\begin{aligned} 50 &= 27 \cdot \underline{1} + 23, \\ 27 &= 23 \cdot \underline{1} + 4, \\ 23 &= 4 \cdot \underline{5} + 3, \\ 4 &= 3 \cdot \underline{1} + 1, \\ 3 &= 1 \cdot \underline{3}, \end{aligned}$$

в котором подчеркнутые числа — неполные частные. Последний ненулевой остаток в алгоритме Евклида — это и есть НОД, поэтому $(50, 27) = 1$.

Далее строим таблицу:

k	-1	0	1	2	3	4
q_k	$-$	1	1	5	1	3
x_k	1	0	1	-1	6	-7
y_k	0	1	-1	2	-11	13

Из данной таблицы видно, что $x = -7$, $y = 13$.

Предложение 1.4. 1. Пусть m — произвольное целое положительное число. Тогда $(am, bm) = (a, b)m$.

2. Пусть d — некоторый общий делитель чисел a и b. Тогда $\left(\dfrac{a}{d}, \dfrac{b}{d}\right) = \dfrac{(a, b)}{d}$.

Доказательство. 1. Умножим на m левые и правые части в цепочке равенств (1.3). Получим равенство $(am, bm) = r_n m$, где $r_n = (a, b)$.

2. Из пункта 1 следует, что:

$$(a, b) = \left(\frac{a}{d}d,\ \frac{b}{d}d\right) = \left(\frac{a}{d},\ \frac{b}{d}\right) d.$$

Из данного равенства следует доказательство пункта 2. □

Предложение 1.5. 1. Если $(a, b) = 1$, тогда $(ac, b) = (c, b)$.

2. Если $(a, b) = 1$ и ac делится на b, тогда c делится на b.

Доказательство. 1. Число (ac, b) является делителем чисел ac и bc. Поэтому по теореме 1.4 данное число делит и $(ac, bc) = c$

(предложение 1.4). Следовательно, (ac, b) делит c и b, поэтому оно делит и (c, b).

Обратно, число (c, b) делит ac и b, поэтому оно делит и (ac, b). Так как числа (ac, b) и (c, b) взаимно делят друг друга и оба положительны, то они равны между собой.

2. Если число ac делится на b, то по предложению 1.2 выполнено равенство $(ac, b) = b$. Из пункта 1 следует, что $(c, b) = b$. Таким образом, c делится на b. □

Предложение 1.6. Пусть каждое из чисел a_1, a_2,. . . ,a_m взаимно просто с каждым из чисел b_1, b_2,. . . ,b_n. Тогда произведение $a_1a_2 \dots a_m$ взаимно просто с произведением $b_1b_2 \dots b_n$.

Доказательство. Из предложения 1.5 следует, что для произвольного $i = 1, 2, \dots, n$ выполнена такая цепочка равенств:

$$(a_1a_2 \dots a_m,\ b_i) = (a_2a_3 \dots a_m,\ b_i) =$$

$$= (a_3 \dots a_m,\ b_i) = \dots = (a_m,\ b_i) = 1.$$

Обозначим $A = a_1a_2 \dots a_m$. Тогда из того же предложения 1.5 следует, что:

$$(b_1b_2 \dots b_n,\ A) = (b_2b_3 \dots b_n,\ A) =$$

$$= (b_3 \dots b_n,\ A) = \dots = (b_n,\ A) = 1.$$ □

1.3. Наименьшее общее кратное

Пусть a_1,. . . ,a_n — некоторые целые числа, причем ни одно из них не равно нулю. Всякое целое число, делящееся одновременно на числа a_1,. . . ,a_n, называется общим кратным данных чисел. Наименьшее из положительных общих кратных чисел a_1,. . . ,a_n называется *наименьшим общим кратным* и обозначается $[a_1, \dots, a_n]$.

Пусть a и b — некоторые ненулевые целые числа и M — некоторое общее кратное данных чисел. В этом случае найдутся два таких целых числа s и t, что $M = as$ и $M = bt$, из чего следует, что $as = bt$. Пусть $d = (a, b)$. Тогда $d\widetilde{a}s = d\widetilde{b}t$, где $a = d\widetilde{a}$, $b = d\widetilde{b}$, причем $(\widetilde{a}, \widetilde{b}) = 1$. Из равенства $\widetilde{a}s = \widetilde{b}t$ и предложения

1.5 следует, что s делится на $\widetilde{b}$ и t делится на $\widetilde{a}$: $s = \widetilde{b}k_1$, $t = \widetilde{a}k_2$. Поэтому из равенства $\widetilde{a}s = \widetilde{b}t$ следует равенство $\widetilde{a}\widetilde{b}k_1 = \widetilde{b}\widetilde{a}k_2$, из которого сразу следует, что $k_1 = k_2 = k$. Поэтому $s = \widetilde{b}k$, $t = \widetilde{a}k$. Таким образом:

$$M = as = a\widetilde{b}k = \frac{ab}{d}k.$$

Обратно, всякое число M такой формы делится одновременно на a и b, поэтому такая форма дает общий вид чисел, кратных a и b, где число k выступает в качестве параметра и может принимать произвольное целое значение. Если a и b положительные целые числа, их наименьшее общее кратное получается при $k = 1$, т.е. $[a, b] = \dfrac{ab}{d}$. Таким образом, верна следующая теорема.

Теорема 1.8. Для наименьшего общего кратного $[a, b]$ любых двух натуральных чисел a и b верны следующие утверждения:

(i) множество общих кратных чисел a и b совпадает со множеством кратных числа $[a, b]$;

(ii) для $[a, b]$ верна следующая формула:

$$[a, b] = \frac{ab}{(a, b)}.$$

Предложение 1.7. 1. Пусть m — произвольное целое положительное число. Тогда $[am, bm] = [a, b]m$.

2. Пусть d — некоторый общий делитель чисел a и b. Тогда $\left[\dfrac{a}{d}, \dfrac{b}{d}\right] = \dfrac{[a, b]}{d}$.

Доказательство следует из теоремы 1.8 и предложения 1.4. □

Теорема 1.9. Для любого набора натуральных чисел a_1, a_2,..., a_n верны следующие утверждения:

(i) $[a_1, a_2, \ldots, a_n] = [[\ldots[[a_1, a_2], a_3], \ldots], a_n]$;

(ii) совокупность общих кратных чисел a_1, a_2,..., a_n совпадает с совокупностью кратных числа $[a_1, a_2, \ldots, a_n]$;

(iii) если числа a_1, a_2,..., a_n попарно взаимно просты, то

$$[a_1, a_2, \ldots, a_n] = a_1 a_2 \ldots a_n.$$

Доказательство. Пункты (i) и (ii) доказываются аналогичным образом, как и теорема 1.5. Покажем пункт (iii). Пусть числа a_1, a_2,..., a_n являются попарно взаимно простыми. Тогда из теоремы 1.8 следует, что $[a_1, a_2] = a_1 a_2$, т.е. база индукции выполнена. Предположим, что для любого $k < n$, где $n \geqslant 3$, выполнено равенство $[a_1, a_2, \ldots, a_k] = a_1 a_2 \ldots a_k$. Тогда, учитывая теорему 1.8 и пункт (i) данной теоремы, имеем:

$$[a_1, a_2, \ldots, a_n] = [[a_1, \ldots, a_{n-1}], a_n] = [a_1 \ldots a_{n-1}, a_n] =$$
$$= \frac{a_1 \ldots a_{n-1} a_n}{(a_1 \ldots a_{n-1}, a_n)} = a_1 \ldots a_{n-1} a_n.$$

Последнее равенство в предыдущей цепочке равенств следует из предложения 1.6. □

1.4. Диофантовы уравнения первой степени

Определение 1.2. *Диофантовым уравнением первой степени* с n неизвестными называется уравнение вида:

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = b, \tag{1.5}$$

где все коэффициенты и неизвестные — целые числа и выполнено условие:

$$a_1^2 + a_2^2 + \ldots + a_n^2 > 0.$$

Определение 1.3. *Решением диофантова уравнения* (1.5) называется множество всех таких $(\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n) \in \mathbb{Z}^n$, удовлетворяющих данному уравнению.

Теорема 1.10. Для любых целых чисел $a_1, \ldots, a_n$, $n \geqslant 2$, существуют такие целые числа $c_1, \ldots, c_n$, что:

$$a_1 c_1 + \ldots + a_n c_n = (a_1, \ldots, a_n).$$

Доказательство. I способ. Пусть не все из чисел $a_1, \ldots, a_n$ равны нулю. Рассмотрим следующее множество:

$$M = \{a_1 x_1 + \ldots + a_n x_n \mid x_1, \ldots, x_n \in \mathbb{Z}\}.$$

Данное множество содержит положительные числа, например:

$$a_1 \cdot sgn(a_1) + a_2 \cdot sgn(a_2) + \ldots + a_n \cdot sgn(a_n).$$

Обозначим через d минимальный положительный элемент множества M. Заметим, что:

$$d = a_1c_1 + \ldots + a_nc_n, \quad c_1, \ldots, c_n \in \mathbb{Z}.$$

Покажем, что $M = \{dx \mid x \in \mathbb{Z}\}$. Пусть $x \in \mathbb{Z}$. Тогда:

$$dx = (a_1c_1 + \ldots + a_nc_n)x = a_1(c_1x) + \ldots + a_n(c_nx) \in M.$$

Поэтому $\{dx \mid x \in \mathbb{Z}\} \subseteq M$. Обратно, пусть $y \in M$:

$$y = a_1y_1 + \ldots + a_ny_n.$$

Разложим число y по модулю числа d (теорема 1.1): $y = dq + r$, $0 \leqslant r < d$. Тогда:

$$r = a_1(y_1 - c_1q) + \ldots + a_n(y_n - c_nq) \in M.$$

Так как d является наименьшим положительным числом в M, то $r = 0$ и $y = dq$. Поэтому $M \subseteq \{dx \mid x \in \mathbb{Z}\}$ и:

$$\{a_1x_1 + \ldots + a_nx_n \mid x_1, \ldots, x_n \in \mathbb{Z}\} = \{dx \mid x \in \mathbb{Z}\}. \qquad (1.6)$$

Осталось показать, что $d = (a_1, \ldots, a_n)$. Из равенства (1.6) следует, что число d делит числа $a_1, \ldots, a_n$, т.е. является их общим делителем. Пусть c — любой другой делитель чисел $a_1, \ldots, a_n$. Тогда из пункта 9 теоремы 1.3 следует, что число c является делителем любой линейной комбинации вида:

$$a_1x_1 + \ldots + a_nx_n, \quad x_1, \ldots, x_n \in \mathbb{Z},$$

одна из которых совпадает с числом $d = a_1c_1 + \ldots + a_nc_n$. Поэтому число c делит и d.

II способ. Применим математическую индукцию по n. База индукции при $n = 1$ и $n = 2$ выполнена (следствие 1.1).

По предположению индукции существует решение уравнения:

$$a_1x_1 + \ldots + a_{n-1}x_{n-1} = (a_1, \ldots, a_{n-1}), \quad n \geqslant 3.$$

Пусть $\widetilde{x}_1, \ldots, \widetilde{x}_{n-1}$ — решение данного уравнения. Учитывая теорему 1.5, рассмотрим уравнение:

$$(a_1, \ldots, a_{n-1})y + a_nx_n = ((a_1, \ldots, a_{n-1}), a_n) = (a_1, \ldots, a_n)$$

относительно неизвестных y и x_n. Пусть $\widetilde{y}$, $\widetilde{x}_n$ — решение последнего уравнения (база индукции). Тогда:

$$\begin{aligned}(a_1,\ldots,a_n) &= (a_1,\ldots,a_{n-1})\widetilde{y}+a_n\widetilde{x}_n = \\ &= (a_1\widetilde{x}_1+\ldots+a_{n-1}\widetilde{x}_{n-1})\widetilde{y}+a_n\widetilde{x}_n = \\ &= a_1(\widetilde{x}_1\widetilde{y})+\ldots+a_{n-1}(\widetilde{x}_{n-1}\widetilde{y})+a_n\widetilde{x}_n.\end{aligned}$$

Поэтому $\widetilde{x}_1\widetilde{y},\ldots,\widetilde{x}_{n-1}\widetilde{y},\widetilde{x}_n$ — решение рассматриваемого уравнения. □

Теорема 1.11. Пусть $d=(a_1,a_2,\ldots,a_n)$. Диофантово уравнение (1.5) имеет решение тогда и только тогда, когда $d|b$. При этом множество решений диофантова уравнения (1.5) либо пусто, либо состоит из бесконечного числа элементов.

Доказательство. Пусть $d|b$. Тогда из теоремы 1.10 следует, что диофантово уравнение (1.5) имеет решение.

Обратно, пусть $\widetilde{x}_1$, $\widetilde{x}_2,\ldots,\widetilde{x}_n$ — некоторое решение диофантово уравнения (1.5) :

$$a_1\widetilde{x}_1+\ldots+a_n\widetilde{x}_n=b.$$

Учитывая что $a_i=\widetilde{a}_i d$, $\widetilde{a}_i\in\mathbb{Z}$, $i=1,\ldots,n$, получаем:

$$d(\widetilde{a}_1\widetilde{x}_1+\ldots+\widetilde{a}_n\widetilde{x}_n)=b,$$

поэтому $d|b$.

Покажем вторую часть теоремы. Пусть множество решений уравнения (1.5) непусто и $\widetilde{x}_1$, $\widetilde{x}_2,\ldots,\widetilde{x}_n$ — одно из его решений. Тогда для любого целого числа t числа вида:

$$\widetilde{x}_1+a_2t,\ \widetilde{x}_2-a_1t,\ \widetilde{x}_3,\ldots,\ \widetilde{x}_n$$

являются решением уравнения (1.5). □

Далее рассматриваются способы нахождения всех решений диофантовых уравнений (1.5) при $n=2$ исходя из частного решения, которое можно найти с помощью расширенного алгоритма Евклида (теорема 1.7).

Теорема 1.12. Пусть для диофантова уравнения:

$$ax+by=c \tag{1.7}$$

выполнены условия: $a \neq 0$, $b \neq 0$, $d = (a, b)$ и $d|c$. Пусть также x_0, y_0 — некоторое частное решение уравнения (1.7). Тогда множество всех решений уравнения (1.7) имеет следующий вид:

$$x_0 - \frac{b}{d}t, \quad y_0 + \frac{a}{d}t, \quad t \in \mathbb{Z}. \tag{1.8}$$

Доказательство. Очевидно, что всякая пара чисел вида (1.8) является решением диофантова уравнения (1.7).

Обратно, пусть $\widetilde{x}$, $\widetilde{y}$ — некоторое решение уравнения (1.7), т.е.:

$$a\widetilde{x} + b\widetilde{y} = c.$$

Вычитая данное равенство из равенства $ax_0 + by_0 = c$, получаем:

$$a(x_0 - \widetilde{x}) = b(\widetilde{y} - y_0).$$

Учитывая, что $d = (a, b)$, приходим к такому равенству:

$$\frac{a}{d}(x_0 - \widetilde{x}) = \frac{b}{d}(\widetilde{y} - y_0).$$

Так как $\left(\frac{a}{d}, \frac{b}{d}\right) = 1$, то из предложения 1.5 следует, что $\frac{a}{d}$ делит $\widetilde{y} - y_0$. Поэтому для некоторого целого t выполнено равенство:

$$\widetilde{y} = y_0 + \frac{a}{d}t.$$

Тогда:

$$a\widetilde{x} = c - b\widetilde{y} = c - b\left(y_0 + \frac{a}{d}t\right) = (c - by_0) - \frac{ab}{d}t = ax_0 - \frac{ab}{d}t,$$

откуда немедленно следует, что:

$$\widetilde{x} = x_0 - \frac{b}{d}t. \qquad \square$$

Следствие 1.2. Пусть x_0, y_0 — некоторое частное решение линейного диофантова уравнения $ax + by = d$, $d = (a, b)$. Тогда множество всех решений уравнения $ax + by = d \cdot q$ имеет следующий вид:

$$x_0 \cdot q - \frac{b}{d}t, \quad y_0 \cdot q + \frac{a}{d}t, \quad t \in \mathbb{Z}.$$

1.5. Системы диофантовых уравнений первой степени

Рассмотрим систему диофантовых уравнений

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ \ldots & \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m, \end{cases} \tag{1.9}$$

где все a_{ij}, b_k — целые числа. Неизвестные $x_1, \ldots, x_n$ — также предполагаются целыми. С системой (1.9) свяжем две матрицы:

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} a_{11} & \ldots & a_{1n} & -b_1 \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} & -b_m \end{pmatrix}.$$

Матрицу A будем называть матрицей системы, B — расширенной матрицей. Пусть $\overline{b}$ — вектор-столбец размерностью m, состоящий из правых частей системы (1.9), $\overline{x} = (x_1, \ldots, x_n)^T$. Тогда в матричном виде систему (1.9) можно записать в виде $A \cdot \overline{x} = \overline{b}$. Если $\overline{b} = \overline{0}$, то данная система называется однородной.

Следуя работе [28], опишем процесс, позволяющий находить все решения системы (1.9) в целых числах.

Составим матрицу $\mathcal{A}$, содержащую $m+n$ строк и $n+1$ столбцов. Первые m строк данной матрицы совпадают со строками матрицы B. Продолжения первых n столбцов матрицы A совпадают со столбцами единичной матрицы E_n порядка n, а продолжение последнего столбца матрицы B состоит из нулей. Таким образом, матрица $\mathcal{A}$ имеет следующий вид:

$$\mathcal{A} = \begin{pmatrix} A & -\overline{b} \\ E_n & \overline{0} \end{pmatrix}.$$

Столбцы матрицы $\mathcal{A}$ в порядке следования слева направо обозначим $[\mathcal{A}]_1, \ldots, [\mathcal{A}]_{n+1}$. Часть матрицы $\mathcal{A}$, образованную столбцами $[\mathcal{A}]_1, \ldots, [\mathcal{A}]_n$, будем называть главной частью матрицы $\mathcal{A}$.

Процесс решения системы (1.9) распадается на два этапа. На первом из них выполняются некоторые преобразования, имеющие целью привести матрицу $\mathcal{A}$ к специальному виду:

$$\widetilde{\mathcal{A}} = \begin{pmatrix} C & -\overline{b} \\ K & \overline{0} \end{pmatrix}, \tag{1.10}$$

где C и K — целочисленные матрицы размером $m \times n$ и $n \times n$ соответственно, при этом матрица C имеет трапециевидную форму:

$$C = \begin{pmatrix} c_{11} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ c_{21} & c_{22} & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{r1} & c_{r2} & c_{r3} & \dots & c_{rr} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & c_{m3} & \dots & c_{mr} & 0 & \dots & 0 \end{pmatrix}, \qquad (1.11)$$

где $c_{11} > 0, \dots, c_{rr} > 0$.

Перечислим преобразования трех типов (допустимые преобразования), которые понадобятся для приведения матрицы $\mathcal{A}$ к виду $\widetilde{\mathcal{A}}$.

1. Перестановка столбцов главной части матрицы $\mathcal{A}$.
2. Изменение знаков на противоположные у всех элементов какого-либо столбца главной части.
3. Умножение первого столбца главной части на некоторое целое число и вычитание результата из другого столбца главной части.

Опишем последовательность выполняемых преобразований подробнее. В ходе действия описанного ниже алгоритма матрица $\mathcal{A}$ будет меняться, но для краткости записи будем обозначать замещающие ее матрицы той же буквой $\mathcal{A}$, а для элементов этих матриц будем использовать те же обозначения a_{ij}.

Заметим, что на любом шаге алгоритма при необходимости к первым m строкам матрицы $\mathcal{A}$ можно применять элементарные преобразования строк.

Алгоритм 1.1 (получение матрицы $\widetilde{\mathcal{A}}$).
I.

1. Изменить главную часть матрицы $\mathcal{A}$ с помощью допустимых преобразований вида 1 и 2 так, чтобы в левом верхнем углу матрицы $\mathcal{A}$ стояло положительное число, наименьшее

среди абсолютных величин всех ненулевых элементов первой строки главной части.

2. Разделить каждое из чисел a_{1j} с остатком на a_{11} (теорема 1.1), т.е. найти представление:

$$a_{1j} = a_{11}q_j + r_j, \quad 0 \leqslant r_j < a_{11}, \quad j = 2, \ldots, n.$$

После этого заменить в матрице $\mathcal{A}$ столбцы $[\mathcal{A}]_2,\ldots,[\mathcal{A}]_n$ столбцами:

$$[\mathcal{A}]_2 - q_2[\mathcal{A}]_1, \ldots, [\mathcal{A}]_n - q_n[\mathcal{A}]_1$$

соответственно.

3. Если среди элементов $a_{12}, \ldots, a_{1n}$ имеются ненулевые элементы, то вернуться к шагу 1.

В результате будет уменьшаться сумма абсолютных величин элементов первой строки главной части матрицы $\mathcal{A}$. Процесс завершится лишь тогда, когда все элементы $a_{12},\ldots,$ a_{1n} станут равными нулю. На этом этапе вычислений определится первый столбец матрицы $\widetilde{\mathcal{A}}$ и $c_{11} = a_{11}$.

II.

Дальнейшая работа происходит со столбцами $[\mathcal{A}]_2,\ldots,[\mathcal{A}]_n$. Так как все первые элементы данных столбцов равны нулю, то после выполнения допустимых преобразований с этими столбцами данные элементы останутся нулевыми. В дальнейшем необходимо выполнять операции шагов 1, 2 и 3 данного алгоритма с матрицей, которая образована столбцами $[\mathcal{A}]_2,\ldots,[\mathcal{A}]_n$, работая с элементами второй строки этой матрицы (напомним, что первая строка нулевая). В результате матрица $\mathcal{A}$ преобразуется к виду $c_{22} = a_{22} > 0$, $a_{23} = \ldots = a_{2n} = 0$.

Продолжение подобных действий со столбцами $[\mathcal{A}]_3,\ldots,[\mathcal{A}]_n$ и т.д. приведет матрицы $\mathcal{A}$ к матрице $\widetilde{\mathcal{A}}$. Столбцы полученной матрицы будем по-прежнему обозначать $[\mathcal{A}]_1,\ldots,[\mathcal{A}]_{n+1}$.

III.

Обозначим через k_1 и s_1 неполное частное и остаток от деления первого элемента столбца $[\mathcal{A}]_{n+1}$ на c_{11} и заменим столбец $[\mathcal{A}]_{n+1}$ на $[\mathcal{A}]_{n+1} - k_1[\mathcal{A}]_1$. Также обозначим через k_2 и s_2 неполное частное и остаток от деления второго элемента нового столбца $[\mathcal{A}]_{n+1}$ на c_{22} и заменим столбец $[\mathcal{A}]_{n+1}$ на $[\mathcal{A}]_{n+1} - k_2[\mathcal{A}]_2$. И так далее, пока столбец $[\mathcal{A}]_{n+1}$ не будет заменен на $[\mathcal{A}]_{n+1} - k_r[\mathcal{A}]_r$. На этом вычисления заканчиваются. Первые элементы столбца $[\mathcal{A}]_{n+1}$ равны $s_1, \ldots, s_r$, причем $0 \leqslant s_j < c_{jj}$, $j = 1, \ldots, r$, а остальные обозначим последовательно $s_{r+1}, \ldots, s_{m+n}$.

Заметим, что после выполнения шага III алгоритма 1.1 первые m элементов последнего столбца матрицы $\widetilde{\mathcal{A}}$ будут иметь вид:

$$\begin{aligned}
&s_1 = -b_1 - k_1c_{11},\\
&s_2 = -b_2 - k_1c_{21} - k_2c_{22},\\
&\ldots\\
&s_r = -b_r - k_1c_{r1} - k_2c_{r2} - \ldots - k_rc_{rr},\\
&s_{r+1} = -b_{r+1} - k_1c_{r+1,1} - k_2c_{r+1,2} - \ldots - k_rc_{r+1,r},\\
&\ldots\\
&s_m = -b_m - k_1c_{m1} - k_2c_{m2} - \ldots - k_rc_{mr}.
\end{aligned} \tag{1.12}$$

Запишем данные равенства в следующем виде:

$$\begin{pmatrix} s_1 \\ \ldots \\ s_m \end{pmatrix} = -\overline{b} - k_1[C]_1 - \ldots - k_r[C]_r.$$

Теорема 1.13. 1. Система уравнений (1.9) разрешима в целых числах тогда и только тогда, когда $s_1 = \ldots = s_m = 0$.

2. Если система (1.9) имеет решение, то, обозначив $q = n - r$ и $\overline{x}_0 \in \mathbb{Z}^n$ — набор с координатами $s_{m+1}, \ldots, s_{m+n}$ в порядке следования, общее решение системы (1.9) будет иметь вид:

$$\overline{x}_0 + t_1[K]_{r+1} + \ldots + t_q[K]_n,$$

где $t_1, \ldots, t_q$ пробегают всевозможные наборы целых чисел, $[K]_j$ — соответствующие столбцы матрицы K из (1.10).

Доказательство. 1. Нетрудно видеть, что каждый столбец главной части матрицы $\widetilde{\mathcal{A}}$ является линейной комбинацией столбцов главной части матрицы $\mathcal{A}$. Поэтому равенства $s_1 = \ldots = s_m = 0$ означают, что столбец $\overline{b}$ является линейной комбинацией столбцов матрицы A с целыми коэффициентами. Таким образом, система (1.9) имеет решение.

Обратно, пусть система уравнений (1.9) разрешима в целых числах. Тогда вектор $\overline{b}$ является линейной комбинацией с целыми коэффициентами столбцов матрицы A.

Нетрудно видеть, что обратное к допустимому преобразованию также является допустимым. Поэтому каждый столбец главной части матрицы $\mathcal{A}$ является линейной комбинацией столбцов главной части матрицы $\widetilde{\mathcal{A}}$. Поэтому вектор $\overline{b}$ является линейной комбинацией с целыми коэффициентами столбцов матрицы C. Пусть $[C]_1, \ldots, [C]_n$ — столбцы матрицы C. Заметим, что все столбцы $[C]_j$, $j > r$, нулевые. Поэтому для некоторых целых чисел $u_1, \ldots, u_r$ выполнено равенство:

$$u_1[C]_1 + \ldots + u_r[C]_r = \overline{b}. \tag{1.13}$$

Из последнего равенства следует, что $u_1 c_{11} = b_1$. Учитывая теорему 1.1 и равенство (1.12), получаем, что $-k_1 = u_1$, $s_1 = 0$. Запишем равенство $u_1 c_{21} + u_2 c_{22} = b_2$ в виде $u_2 c_{22} = b_2 + k_1 c_{21}$. Тогда из (1.12) и теоремы 1.1 следует, что $-k_2 = u_2$ и $s_2 = 0$. Продолжая данные рассуждения, находим $s_1 = \ldots = s_r = 0$, $-k_j = u_j$, $j = 1, \ldots, r$. Из равенства (1.13) следует, что:

$$b_j = u_1 c_{j1} + \ldots + u_r c_{jr} = -k_1 c_{j1} - \ldots - k_r c_{jr}, \quad j = r+1, \ldots, m.$$

Поэтому из (1.12) следует, что $s_{r+1} = \ldots = s_m = 0$.

2. Заметим, что любая допустимая операция со столбцами матрицы равносильна умножению этой матрицы на некоторую целочисленную матрицу с определителем ± 1 с правой стороны. Изменение знака у всех элементов столбца с номером j соответствует умножению с правой стороны на диагональную матрицу, у которой все элементы главной диагонали равны 1, за исключением элемента с индексами (j, j), который равен -1.

Перестановка столбцов с номерами i и j соответствует умножению на матрицу, которая получена из единичной матрицы путем перестановки столбцов с i-м и j-м номерами. Вычитание первого столбца, умноженного на целое q, из столбца с номером j соответствует умножению на матрицу, которая получена из единичной матрицы путем добавления элемента со значением $-q$ на позицию $(1, j)$.

Таким образом, выполнено следующее равенство:

$$\begin{pmatrix} C \\ K \end{pmatrix} = \begin{pmatrix} A \\ E_n \end{pmatrix} \cdot S,$$

где S — некоторая целочисленная матрица размером $n \times n$ с определителем ± 1. Тогда $C = AS$, $K = E_n S$ и $C = AK$.

Заметим, что после выполнения шага III алгоритма 1.1 последние n элементов последнего столбца матрицы $\widetilde{\mathcal{A}}$ будут иметь вид:

$$\begin{pmatrix} s_{m+1} \\ \dots \\ s_{m+n} \end{pmatrix} = -k_1[K]_1 - \ldots - k_r[K]_r.$$

Пусть система (1.9) имеет решение. Тогда:

$$A \cdot \overline{x}_0 = A \cdot (-k_1[K]_1 - \ldots - k_r[K]_r) = -k_1[C]_1 - \ldots - k_r[C]_r = \overline{b}.$$

Поэтому $\overline{x}_0$ — решение. Так как:

$$A \cdot [K]_j = [C]_j = \overline{0}, \quad j = r+1, \ldots, n,$$

то:

$$A \cdot (\overline{x}_0 + t_1[K]_{r+1} + \ldots + t_q[K]_n) = \overline{b}$$

и $\overline{x}_0 + t_1[K]_{r+1} + \ldots + t_q[K]_n$ является решением системы (1.9) для любых целых чисел $t_1, \ldots, t_q$.

С другой стороны, пусть $\overline{x} \in \mathbb{Z}^n$ — некоторое решение системы (1.9). Тогда $A(\overline{x} - \overline{x}_0) = \overline{0}$. Так как $K = S$, то определитель матрицы K равен ± 1 и обратная к K матрица является целочисленной. Обозначим $\overline{y} = K^{-1}(\overline{x} - \overline{x}_0) \in \mathbb{Z}^n$. Имеем:

$$C\overline{y} = AK\overline{y} = A(\overline{x} - \overline{x}_0) = \overline{0}.$$

Учитывая строение матрицы C (1.11), получаем, что первые r координат вектора $\overline{y}$ равны нулю. Поэтому:

$$\overline{x} - \overline{x}_0 = K\overline{y} = y_{r+1}[K]_{r+1} + \ldots + y_n[K]_n. \qquad \square$$

Пример 1.2. Найти общее решение уравнения:

$$7x_1 + 6x_2 + 8x_3 = 1.$$

Выполняя допустимые преобразования, находим:

$$\left(\begin{array}{ccc|c} 7 & 6 & 8 & -1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 6 & 7 & 8 & -1 \\ \hline 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 6 & 1 & 2 & -1 \\ \hline 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \sim$$

$$\sim \left(\begin{array}{ccc|c} 1 & 6 & 2 & -1 \\ \hline 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ \hline 1 & -6 & -2 & 1 \\ -1 & 7 & 1 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Данное уравнение разрешимо и его общее решение имеет вид:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + t_1 \begin{pmatrix} -6 \\ 7 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \quad t_1, t_2 \in \mathbb{Z}.$$

1.6. Простые числа

Определение 1.4. Натуральное число $p > 1$ называется простым, если оно имеет ровно два положительных делителя, а именно 1 и p. В противном случае число p, $p > 1$, называется составным.

Предложение 1.8. Для любого натурального числа $n > 1$ наименьший отличный от единицы делитель всегда есть простое число.

Доказательство. Пусть q — наименьший отличный от единицы делитель числа $n > 1$. Предположим, что q является составным, т.е. число q имеет делитель $\widetilde{q}$, $1 < \widetilde{q} < q$. Тогда по

теореме 1.3 число n делится на $\widetilde{q}$. Противоречие с тем, что q является наименьшим отличным от единицы делителем числа n. $\square$

Предложение 1.9. Наименьший отличный от единицы делитель составного числа n не превосходит $\sqrt{n}$.

Доказательство. Пусть q — наименьший отличный от единицы делитель составного числа n. Тогда $n = qd$, причем:

$$1 < q \leqslant d < n.$$

Поэтому $n = qd \geqslant q^2$ и $q \leqslant \sqrt{n}$. $\square$

Следствие 1.3. Если натуральное число $n > 1$ не делится ни на одно простое число, не превосходящее $\sqrt{n}$, то оно простое.

На данном утверждении основан *метод пробных делений* проверки числа a на простоту. При этом перебираются все числа $d = 2, 3, \ldots, [\sqrt{a}]$ и проверяется, делится ли число a на d. Если среди данного набора делитель не будет найден, то число a является простым. Данный метод можно немного усовершенствовать следующим образом. Пусть m — некоторое натуральное число, причем $m > 1$. Пусть также $i_1, \ldots, i_k$ — все числа, взаимно простые с m, из множества $1, \ldots, m-1$. Пусть p — некоторое простое число. Тогда при разложении числа p по модулю числа m ($p = qm + r$) остаток r будет принадлежать множеству $\{i_1, \ldots, i_k\}$. Поэтому при проверке числа a на простоту среди чисел $d = 2, 3, \ldots, [\sqrt{a}]$ можно рассматривать лишь те, которые имеют вид $qm + i$, $q = 1, 2, \ldots$, $i \in \{i_1, \ldots, i_k\}$, при этом предварительно проверив все простые числа, не превышающие числа m.

Например, если $m = 6$, то достаточно перебирать числа вида $6q + 1$ и $6q + 5$, $q = 1, 2, \ldots$, предварительно проверив, не являются ли числа 2 и 3 делителями числа a. Видно, что в этом случае из чисел $\{2, 3, \ldots, [\sqrt{a}]\}$ отбрасывается примерно $2/3$.

Теорема 1.14. (i) Если в наборе чисел $2, 3, \ldots, N$ вычеркнуть все числа, кратные первым s простым числам $2, 3, \ldots, p_s$

(включая и сами $2, 3, \ldots, p_s$), то первое невычеркнутое число (наименьшее из невычеркнутых) будет простым числом.

(*ii*) Если вычеркнуть все числа, кратные первым s простым числам $2, 3, \ldots, p_s$, где s такое, что $p_s \leqslant \sqrt{N} < p_{s+1}$, кроме самих $2, 3, \ldots, p_s$, то останутся только простые числа, причем данный набор простых чисел совпадает с набором всех простых чисел, не превосходящих числа N.

Доказательство следует из предложений 1.8 и 1.9. □

Из теоремы 1.14 следует алгоритм 1.2.

Алгоритм 1.2 (**решето Эратосфена**).
Вход: натуральное число N.

Выход: все простые числа $p_1 < p_2 < \ldots < p_s$ в диапазоне от 2 до N.

1. Выпишем все целые числа 2,3,4,5,. . . ,N. Положим $p_1 = 2$ и начиная с $4 = p_1^2$ будем вычеркивать числа, двигаясь с шагом 2. Заметим, что на данном шаге вычеркиваются все четные числа, кроме числа 2.

2. Пусть $k \geqslant 2$ и определены числа p_1,. . . , p_{k-1}. Обозначим через p_k первое невычеркнутое число, следующее за p_{k-1}. Если $p_k^2 > N$, то оставшиеся невычеркнутые числа (следующие за p_k) обозначим последовательно через p_{k+1}, p_{k+2},. . . . На этом алгоритм свою работу завершает.

3. Если $p_k^2 \leqslant N$, то вычеркиваем числа, начиная с p_k^2 и двигаясь до N с шагом $2p_k$ (здесь учтен первый шаг алгоритма). Вычеркнутые ранее числа принимаются в учет, но не вычеркиваются еще раз. По завершении этого шага алгоритм увеличивает индекс k на единицу и переходит к шагу 2.

Теорема 1.15 (**Евклид**). Простых чисел бесконечно много.

Доказательство. Предположим, что множество простых чисел конечно и состоит из различных чисел p_1, p_2,. . . ,p_n. Рассмотрим натуральное число $a = p_1 p_2 \ldots p_n + 1$. Заметим, что число a не делится ни на одно из простых чисел p_1, p_2,. . . ,p_n, так как при делении на данные числа в остатке получается число 1.

Таким образом, число a не делится ни на одно простое число, поэтому по предложению 1.8 $a = 1$. Но из определения числа a видно, что $a > 1$. Противоречие. □

Предложение 1.10. Для любого натурального числа n в натуральном ряду найдется последовательность из n подряд идущих составных элементов.

Доказательство. В этом случае достаточно рассмотреть последовательность чисел:

$$(n+1)! + 2,\ (n+1)! + 3, \ldots,\ (n+1)! + n + 1,$$

в которой первый элемент делится на 2, второй — на 3,. . . , n-й — на $n + 1$. □

Предложение 1.11. Пусть a — произвольное целое число и p — произвольное простое число. Тогда либо a делится на p, либо числа a и p взаимно просты.

Доказательство. Предположим, что числа a и p не являются взаимно простыми. Так как p делится на (a, p) и $(a, p) > 1$, то $(a, p) = p$, поэтому a делится на p. □

Предложение 1.12. Пусть a_1, a_2,. . . ,a_n — произвольные целые числа и p — произвольное простое число, причем произведение $a_1 a_2 \ldots a_n$ делится на p. Тогда хотя бы одно из данных чисел делится на p.

Доказательство. Каждое из чисел a_1, a_2,. . . ,a_n либо взаимно просто с p, либо делится на p (предложение 1.11). Предположим, что все числа a_1, a_2,. . . ,a_n взаимно просты с p. Тогда из предложения 1.6 следует, что произведение $a_1 a_2 \ldots a_n$ взаимно просто с p. Противоречие. Поэтому хотя бы одно из данных чисел делится на p. □

Теорема 1.16 (**основная теорема арифметики**). Каждое натуральное число $a > 1$ представимо в виде произведения простых чисел, причем данное разложение единственно (с точностью до порядка следования сомножителей).

Доказательство. Пусть a — некоторое натуральное число и $a > 1$. Обозначим через p_1 наименьший простой делитель числа a: $a = p_1 a_1$. Если $a_1 > 1$, тогда обозначим через p_2 наименьший простой делитель числа a_1: $a_1 = p_2 a_2$. Если же $a_2 > 1$, то аналогично получаем $a_2 = p_3 a_3$ и т.д. Так как $a > a_1 > a_2 > \ldots > 0$, то данный процесс остановится на некотором $a_n = 1$, при этом $a_{n-1} = p_n$. Получаем $a = p_1 p_2 \ldots p_n$.

Покажем единственность такого разложения. Предположим, что для того же самого a найдется и другое разложение: $a = q_1 q_2 \ldots q_m$. Тогда:

$$p_1 p_2 \ldots p_n = q_1 q_2 \ldots q_m.$$

Так как правая часть данного равенства делится на q_1, то по предложению 1.12 хотя бы одно из чисел p_1, p_2,. . . ,p_n делится на q_1. Не ограничивая общности, пусть это будет p_1. Так как p_1 простое число, то $p_1 = q_1$. Сократив обе части равенства на $p_1 = q_1$, получим $p_2 p_3 \ldots p_n = q_2 q_3 \ldots q_m$. Рассуждая аналогично, получим $p_2 = q_2$,. . . ,$p_k = q_k$, где $k = \min\{n, m\}$. Если $n = m$, то все доказано. Если же, например, $m > n$, то $k = n$ и остается равенство $q_{n+1} \ldots q_m = 1$. Но все q_1,. . . ,q_m больше 1, поэтому приходим к противоречию. То же самое будет и при $n > m$. □

Определение 1.5. Каноническим разложением натурального числа $a > 1$ называется представление числа a в виде:

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n},$$

где $p_1 < \ldots < p_n$ — попарно различные простые числа, α_1,. . . , α_n — натуральные числа.

Из предложения 1.12 непосредственно следуют такие утверждения.

Теорема 1.17. Пусть a — некоторое натуральное число, причем $a = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$ — каноническое разложение данного числа. Тогда если d — некоторый делитель числа a, то d имеет следующий вид:

$$d = p_1^{\beta_1} p_2^{\beta_2} \ldots p_n^{\beta_n}, \text{ причем:}$$

$$0 \leqslant \beta_1 \leqslant \alpha_1,\ 0 \leqslant \beta_2 \leqslant \alpha_2, \ldots,\ 0 \leqslant \beta_n \leqslant \alpha_n.$$

Теорема 1.18. Пусть натуральные числа a_1, a_2,..., a_m представимы в следующем виде:

$$a_1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n},\ a_2 = p_1^{\beta_1} p_2^{\beta_2} \dots p_n^{\beta_n}, \dots,\ a_m = p_1^{\gamma_1} p_2^{\gamma_2} \dots p_n^{\gamma_n},$$

где p_1, p_2,..., p_n — попарно различные простые числа. Тогда:

$$(a_1, a_2, \dots, a_m) = p_1^{\min\{\alpha_1,\beta_1,\dots,\gamma_1\}} p_2^{\min\{\alpha_2,\beta_2,\dots,\gamma_2\}} \dots p_n^{\min\{\alpha_n,\beta_n,\dots,\gamma_n\}},$$

$$[a_1, a_2, \dots, a_m] = p_1^{\max\{\alpha_1,\beta_1,\dots,\gamma_1\}} p_2^{\max\{\alpha_2,\beta_2,\dots,\gamma_2\}} \dots p_n^{\max\{\alpha_n,\beta_n,\dots,\gamma_n\}}.$$

Следствие 1.4. Пусть $a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_m^{\alpha_m}$ и $b = q_1^{\beta_1} q_2^{\beta_2} \dots q_n^{\beta_n}$ — каноническое разложение положительных целых чисел a и b. $(a, b) = 1$ тогда и только тогда, когда:

$$\{p_1,\ p_2, \dots,\ p_m\} \cap \{q_1,\ q_2, \dots,\ q_n\} = \varnothing.$$

Следствие 1.5. Для любых натуральных чисел $b, a_1, \dots, a_n$ выполнено:

$$(a_1 \dots a_n, b) \mid (a_1, b) \dots (a_n, b).$$

Если числа $a_1, \dots, a_n$ попарно взаимно просты, то выполнено равенство:

$$(a_1 \dots a_n, b) = (a_1, b) \dots (a_n, b).$$

Центральное место в проблеме распределения простых чисел занимает описание числовой функции $\pi : (1, +\infty) \to \mathbb{N}$, где $\pi(x)$ — число простых чисел на отрезке $[1; x]$. Чебышев доказал, что функция π аппроксимируется функцией $\dfrac{x}{\ln\, x}$.

Теорема 1.19 (Чебышев). Существуют такие положительные числа $a < 1 < b$, что для любого $x \geqslant 2$ выполнены неравенства:

$$a\frac{x}{\ln\, x} < \pi(x) < b\frac{x}{\ln\, x}.$$

Доказательство можно найти, например, в [7].

Следствие 1.6 (постулат Бертрана). Для любого $n > 1$ между n и $2n - 2$ обязательно найдется хотя бы одно простое число.

Теорема 1.20 (**Дирихле**)**.** Для любых натуральных взаимно простых чисел a и b в последовательности $ax + b$, $x \in \mathbb{N}$, содержится бесконечно много простых чисел.

Доказательство можно найти, например, в [11].

1.7. Разложение $n!$ на простые множители

Определение 1.6. Целой частью действительного числа x (обозначается $[x]$) называется наибольшее целое число, не превосходящее x, т.е. $[x] \leqslant x < [x] + 1$.

Сложив неравенства $[x] \leqslant x$ и $[y] \leqslant y$, получим $[x]+[y] \leqslant x+y$. Поэтому:

$$[x] + [y] \leqslant [x + y].$$

Это неравенство справедливо и при любом количестве слагаемых, что легко доказывается с помощью математической индукции.

Также из определения целой части числа следует, что для любого целого n выполнено $[x + n] = [x] + n$.

Предложение 1.13. Пусть x — положительное действительное число, d — положительное целое. Тогда число положительных целых чисел, не превосходящих x и делящихся на d, равно $\left[\dfrac{x}{d}\right]$.

Доказательство. Рассмотрим положительные целые числа, кратные d и не превосходящие x. Пусть наибольшее из них будет sd. Число таких чисел:

$$d,\ 2d,\ 3d, \ldots,\ sd$$

равно s. Из $sd \leqslant x < (s + 1)d$ следует, что $s \leqslant \dfrac{x}{d} < s + 1$, т.е. $s = \left[\dfrac{x}{d}\right]$. □

Предложение 1.14. Для любого положительного действительного числа x и положительного целого d выполнено равенство $\left[\dfrac{[x]}{d}\right] = \left[\dfrac{x}{d}\right]$.

Доказательство. Так как между числами $[x]$ и x нет целых чисел, то количество целых чисел из отрезка $[1, [x]]$, кратных d, равно количеству целых чисел из отрезка $[1, x]$, кратных d. Поэтому из предложения 1.13 получаем требуемое равенство. □

Предложение 1.15. Пусть p — простое число, n — натуральное. Тогда для показателя s наивысшей степени p, делящей $n!$, выполнено равенство

$$s = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \dots, \tag{1.14}$$

т.е. при s, равном сумме (1.14), $p^s \mid n!$, но $p^{s+1} \nmid n!$.

Доказательство. Применим математическую индукцию по n при фиксированном p. База индукции при $n < p$ выполнена, так как, учитывая предложение 1.12, $s = 0$. При этом все слагаемые в (1.14) также равны нулю.

Предположим, что утверждение верно для любого $k < n$. Покажем справедливость утверждения для случая $k = n$. В этом случае $n \geqslant p$. Среди множителей $1, 2, \dots, n$ числа $n!$ количество чисел, делящихся на p, равно $\left[\dfrac{n}{p}\right]$ (предложение 1.13):

$$n! = \dots \cdot p \cdot \dots \cdot 2p \cdot \dots \cdot 3p \cdot \dots \cdot \left[\frac{n}{p}\right] p \cdot \dots =$$

$$= \left[\frac{n}{p}\right]! p^{\left[\frac{n}{p}\right]} m, \quad p \nmid m. \tag{1.15}$$

Так как $1 \leqslant \left[\dfrac{n}{p}\right] < n$, то к числу $\left[\dfrac{n}{p}\right]!$ можно применить предположение индукции. Поэтому показатель наивысшей степени

p, делящей $\left[\frac{n}{p}\right]!$, равен:

$$\left[\frac{\left[\frac{n}{p}\right]}{p}\right] + \left[\frac{\left[\frac{n}{p}\right]}{p^2}\right] = \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \dots .$$

В последнем равенстве использовалось предложение 1.14. Осталось учесть равенство (1.15). □

Замечание 1.2. Так как для некоторого натурального k выполнено $p^k > n$, то $\left[\frac{n}{p^k}\right] = 0$. Поэтому начиная с некоторого номера, все члены ряда (1.14) равны нулю и данный ряд является конечным.

Предложение 1.16. Пусть p — простое число, n — натуральное. Тогда для показателя s наивысшей степени p, делящей $n!$, выполнено неравенство $s < n$. Если же $p > 2$ и $n > 1$, то $s + 1 < n$.

Доказательство. Учитывая равенство (1.14) и замечание 1.2, получаем:

$$s = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \ldots < \frac{n}{p} + \frac{n}{p^2} + \ldots =$$

$$= \frac{n}{p}\left(1 + \frac{1}{p} + \frac{1}{p^2} + \ldots\right) = \frac{n}{p-1} \leqslant n.$$

Пусть $p > 2$ и $n > 1$. Тогда $s + 1 < n$ при $s = 0$. Пусть $s > 0$. Тогда из неравенств $s < \frac{n}{p-1} \leqslant \frac{n}{2}$ следует, что $2s < n$. Поэтому $s + 1 \leqslant s + s < n$. □

Предложение 1.15 дает возможность находить каноническое разложение числа $n!$:

$$n! = \prod_p p^{\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \ldots},$$

где в произведении участвуют все простые числа в диапазоне от 2 до n.

Пример 1.3. Найдем каноническое разложение числа 15!. Простые числа, участвующие в данном разложении, это 2, 3, 5, 7, 11, 13. Для каждого из них найдем степень:

$$2: \ \left[\frac{15}{2}\right]+\left[\frac{15}{4}\right]+\left[\frac{15}{8}\right]=11, \quad 3: \ \left[\frac{15}{3}\right]+\left[\frac{15}{9}\right]=6,$$

$$5: \ \left[\frac{15}{5}\right]=3, \quad 7: \ \left[\frac{15}{7}\right]=2, \quad 11: \ \left[\frac{15}{11}\right]=1, \quad 13: \ \left[\frac{15}{13}\right]=1.$$

Таким образом, $15!=2^{11}\cdot 3^6\cdot 5^3\cdot 7^2\cdot 11\cdot 13$.

1.8. Цепные дроби

1.8.1. Конечные цепные дроби

Определение 1.7. *Конечной непрерывной дробью* называется рациональное число, записанное в виде:

$$a_0+\cfrac{b_1}{a_1+\cfrac{b_2}{a_2+}}\ddots\cfrac{}{+\cfrac{b_{n-1}}{a_{n-1}+\cfrac{b_n}{a_n}}},$$

где a_0, a_1,..., a_n, b_1, b_2,..., b_n — целые числа, причем ни один из знаменателей не равен нулю.

Если $b_1=b_2=\ldots=b_n=1$, $a_i\geqslant 1$ для всех $i=1,2,\ldots,n-1$ и $a_n>1$, то такую непрерывную дробь называют цепной дробью (или обыкновенной непрерывной дробью).

Определение 1.8. *Конечной цепной дробью* называется ра-

циональное число, записанное в виде:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}, \tag{1.16}$$

где $a_0 \in \mathbb{Z}$, $a_1, \ldots, a_n \in \mathbb{N}$, причем $a_n > 1$.

Для краткости записи цепную дробь (1.16) будем иногда записывать в таком виде:

$$[a_0; a_1, \ldots, a_n].$$

Предложение 1.17. Для любой конечной цепной дроби при $n \geqslant 1$ выполнено следующее двойное неравенство:

$$a_0 < [a_0; a_1, \ldots, a_n] < a_0 + 1,$$

в частности, целая часть от цепной дроби $[a_0; a_1, \ldots, a_n]$ равна a_0.

Доказательство. Применим индукцию по n. Пусть $n = 1$. Тогда $a_1 > 1$ и $0 < 1/a_1 < 1$. Поэтому $a_0 < a_0 + 1/a_1 < a_0 + 1$. Предположим, что утверждение верно для любого $k < n$ при $n \geqslant 2$.

Рассмотрим случай $k = n$. Тогда из предположения индукции следует, что:

$$1 \leqslant a_1 < [a_1; a_2, \ldots, a_n] < a_1 + 1,$$

поэтому:

$$0 < \frac{1}{[a_1; a_2, \ldots, a_n]} < 1.$$

Из последнего двойного неравенства немедленно следует справедливость утверждения, так как:

$$[a_0; a_1, \ldots, a_n] = a_0 + \frac{1}{[a_1; a_2, \ldots, a_n]}. \qquad \square$$

Теорема 1.21. Любое рациональное число равно некоторой конечной цепной дроби. Причем для каждого такого числа существует одна и только одна конечная цепная дробь, равная данному числу.

Доказательство. Любое рациональное число можно представить в виде $\dfrac{a}{b}$, где a и b — целые числа, причем $b \geqslant 1$. Применяя алгоритм Евклида, получаем такую цепочку равенств:

$$
\begin{aligned}
&a = bq_0 + r_1: && \frac{a}{b} = q_0 + \frac{1}{\dfrac{b}{r_1}},\\
&b = r_1q_1 + r_2: && \frac{b}{r_1} = q_1 + \frac{1}{\dfrac{r_1}{r_2}},\\
&\dots\\
&r_{n-2} = r_{n-1}q_{n-1} + r_n: && \frac{r_{n-2}}{r_{n-1}} = q_{n-1} + \frac{1}{\dfrac{r_{n-1}}{r_n}},\\
&r_{n-1} = r_nq_n: && \frac{r_{n-1}}{r_n} = q_n.
\end{aligned}
$$

Поэтому:

$$
\frac{a}{b} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ddots + \cfrac{1}{q_{n-1} + \cfrac{1}{q_n}}}}.
$$

Покажем единственность представления любого рационального числа в виде конечной цепной дроби. Предположим, что для числа $\dfrac{a}{b}$ существует еще одно представление в виде конечной цепной дроби:

$$
\frac{a}{b} = [q_0; q_1, \dots, q_n] = [\widetilde{q}_0; \widetilde{q}_1, \dots, \widetilde{q}_t], \quad t \geqslant n. \tag{1.17}
$$

Учитывая предложения 1.17 и взяв целую часть от обеих частей равенства (1.17), получим, что $q_0 = \widetilde{q}_0$. Из этого равенства следует, что будут равны такие цепные дроби:

$$[q_1; q_2, \ldots, q_n] = [\widetilde{q}_1; \widetilde{q}_2, \ldots, \widetilde{q}_t].$$

Взяв от обеих частей данного равенства целую часть, получим, что $q_1 = \widetilde{q}_1$. Продолжая данный процесс, придем к такому равенству:

$$q_n = [\widetilde{q}_n; \widetilde{q}_{n+1}, \ldots, \widetilde{q}_t].$$

Взяв в очередной раз от обеих частей данного равенства целую часть, получим, что $q_n = \widetilde{q}_n$, поэтому:

$$0 = [\widetilde{q}_{n+1}; \widetilde{q}_{n+2}, \ldots, \widetilde{q}_t].$$

Если предположить, что $t > n$, получим, что правая часть данного равенства строго больше 0, поэтому придем к противоречию. Таким образом, $t = n$ и две конечные цепные дроби в равенстве (1.17) равны между собой. □

Для данной цепной дроби:

$$[a_0; a_1, \ldots, a_n] \tag{1.18}$$

рассмотрим так называемые *подходящие дроби*:

$$A_0 = a_0, \quad A_1 = [a_0; a_1], \quad A_2 = [a_0; a_1, a_2], \ldots, A_n = [a_0; a_1, \ldots, a_n].$$

Определение 1.9. *k-ой подходящей дробью* $(0 \leqslant k \leqslant n)$ к конечной цепной дроби (1.18) будем называть величину:

$$A_k = [a_0; a_1, \ldots, a_k].$$

Рассмотрим две последовательности чисел:

$$P_{-1}, P_0, \ P_1, \ \ldots, \ P_n \ \text{ и } Q_{-1}, Q_0, \ Q_1, \ \ldots, \ Q_n,$$

которые определим такими рекуррентными соотношениями при $1 \leqslant k \leqslant n$:

$$\begin{aligned} P_k &= P_{k-2} + P_{k-1}a_k, \\ Q_k &= Q_{k-2} + Q_{k-1}a_k \end{aligned} \tag{1.19}$$

и начальными условиями:

$$\begin{aligned} P_{-1} &= 1, \quad P_0 = a_0, \\ Q_{-1} &= 0, \quad Q_0 = 1. \end{aligned} \tag{1.20}$$

Очевидно, что величины $P_1,\ P_2,\ \ldots,\ P_n$ и $Q_1,\ Q_2,\ \ldots,\ Q_n$ однозначно определяются при заданных $a_0,\ a_1,\ldots,\ a_n$.

Следующая теорема дает простой способ вычисления k-ой подходящей дроби.

Предложение 1.18. Пусть $a_0,\ a_1,\ldots,\ a_n$ — элементы цепной дроби (1.18). Тогда для любого $k = 0, 1, \ldots, n$ значение k-ой подходящей дроби A_k равно $\dfrac{P_k}{Q_k}$, где величины P_k и Q_k определяются рекуррентными соотношениями (1.19) с начальными условиями (1.20).

Доказательство. Доказательство проведем методом математической индукции по n. При $n = 0$ и $n = 1$ имеем:

$$\frac{P_0}{Q_0} = \frac{a_0}{1} = a_0 = A_0,$$

$$\frac{P_1}{Q_1} = \frac{a_0 a_1 + 1}{a_1} = a_0 + \frac{1}{a_1} = A_1.$$

Предположим, что утверждение верно для k, $0 \leqslant k \leqslant n$, т.е.:

$$\frac{P_k}{Q_k} = \frac{P_{k-2} + P_{k-1} a_k}{Q_{k-2} + Q_{k-1} a_k} = A_k.$$

Заметим, что если в k-ой подходящей дроби A_k значение a_k заменить на $a_k + \dfrac{1}{a_{k+1}}$, то получим A_{k+1}. Поэтому:

$$A_{k+1} = \frac{P_{k-2} + P_{k-1}\left(a_k + \dfrac{1}{a_{k+1}}\right)}{Q_{k-2} + Q_{k-1}\left(a_k + \dfrac{1}{a_{k+1}}\right)} = \frac{P_{k-2} + P_{k-1} a_k + \dfrac{P_{k-1}}{a_{k+1}}}{Q_{k-2} + Q_{k-1} a_k + \dfrac{Q_{k-1}}{a_{k+1}}} =$$

$$= \frac{P_k + \dfrac{P_{k-1}}{a_{k+1}}}{Q_k + \dfrac{Q_{k-1}}{a_{k+1}}} = \frac{P_k a_{k+1} + P_{k-1}}{Q_k a_{k+1} + Q_{k-1}} = \frac{P_{k+1}}{Q_{k+1}}. \qquad \square$$

Последовательное вычисление значений P_k и Q_k, определенных соотношениями (1.19) и (1.20), очень удобно производить по следующей схеме:

k	-1	0	1	...	n
a_k	-	a_0	a_1	...	a_n
P_k	1	a_0	$P_{-1} + P_0 a_1$	...	$P_{n-2} + P_{n-1} a_n$
Q_k	0	1	$Q_{-1} + Q_0 a_1$	...	$Q_{n-2} + Q_{n-1} a_n$

Пример 1.4. Найдем подходящие дроби к цепной дроби: $[2; 3, 1, 2, 2]$.

k	-1	0	1	2	3	4
a_k	-	2	3	1	2	2
P_k	1	2	7	9	25	59
Q_k	0	1	3	4	11	26

Таким образом:

$$\frac{P_0}{Q_0} = \frac{2}{1}, \ \frac{P_1}{Q_1} = \frac{7}{3}, \ \frac{P_2}{Q_2} = \frac{9}{4}, \ \frac{P_3}{Q_3} = \frac{25}{11}, \ \frac{P_4}{Q_4} = \frac{59}{26}.$$

Заметим, что из теоремы 1.21 непосредственно следует, что:

$$\frac{59}{26} = [2; 3, 1, 2, 2].$$

Предложение 1.19. При $k = 0, 1, \ldots, n$ выполняется следующее равенство:

$$P_{k-1} Q_k - P_k Q_{k-1} = (-1)^k. \tag{1.21}$$

Доказательство. Доказательство проведем методом математической индукции. При $k = 0$ равенство очевидно. При $k = 1$:

$$P_0 Q_1 - P_1 Q_0 = a_0 a_1 - (a_0 a_1 + 1) = -1.$$

Предположим, что при некотором k, $1 \leqslant k \leqslant n-1$, выполняется равенство (1.21). Тогда:

$$P_k Q_{k+1} - P_{k+1} Q_k = P_k (Q_{k-1} + Q_k a_{k+1}) - (P_{k-1} + P_k a_{k+1}) Q_k =$$

$$= P_k Q_{k-1} - P_{k-1} Q_k = -(P_{k-1} Q_k - P_k Q_{k-1}) = (-1)^{k+1}. \quad \square$$

Предложение 1.20. Для любого $k = 0, 1, \ldots, n$ пары чисел P_k и Q_k являются взаимно простыми.

Доказательство проходит методом все той же индукции. При $k = 0$ $P_0 = a_0$, $Q_0 = 1$, поэтому $(P_0, Q_0) = 1$. Пусть утверждение теоремы верно для некоторого $k-1$, $0 \leqslant k-1 \leqslant n-1$. Покажем, что числа P_k и Q_k взаимно просты.

Действительно, пусть $d = (P_k, Q_k)$. Из равенства (1.21) следует, что d делит число $(-1)^k$. Следовательно, $d = 1$. □

Замечание 1.3. Если рациональное число a/b разложить в цепную дробь, то последняя подходящая дробь P_n/Q_n представляет собой несократимую дробь, равную a/b.

Предложение 1.21. Для любого $k = 1, \ldots, n$ выполнены следующие равенства:

$$\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k} = \frac{(-1)^k}{Q_{k-1}Q_k},$$

$$\left|\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k}\right| = \frac{1}{Q_{k-1}Q_k}.$$

Доказательство следует из предложения 1.19. □

Предложение 1.22. Последовательность $Q_1, \ldots, Q_n$ строго возрастает:

$$1 = Q_0 \leqslant Q_1 < Q_2 < \ldots < Q_n.$$

Доказательство. Применим индукцию по n. При $n = 1$: $Q_1 = a_1 \geqslant 1 = Q_0$. При $n = 2 : Q_2 = Q_1a_2 + Q_0 \geqslant Q_1 \cdot 1 + 1 > Q_1$. Предположим, что утверждение верно при $k < n$. Тогда:

$$Q_n = Q_{n-1}a_n + Q_{n-2} \geqslant Q_{n-1} \cdot 1 + 1 > Q_{n-1}.$$ □

Предложение 1.23. Последовательность $\{P_k\}_{k \geqslant -1}$ положительной цепной дроби, начиная с $k = 0$, строго возрастает:

$$P_0 < P_1 < \ldots < P_n.$$

Доказательство аналогично доказательству предложения 1.22, так как если цепная дробь положительна, то $a_0 \geqslant 0$ и $P_0 = a_0 < 1 + a_0a_1 = P_1$. □

Предложение 1.24. Для любого $k = 1, \ldots, n$ выполнены следующие равенства:

$$P_{k-2}Q_k - P_kQ_{k-2} = (-1)^{k-1}a_k.$$

Доказательство. Учитывая предложение 1.19, получаем:

$$P_{k-2}Q_k - P_kQ_{k-2} = P_{k-2}(Q_{k-2} + Q_{k-1}a_k) - (P_{k-2} + P_{k-1}a_k)Q_{k-2} =$$

$$= (P_{k-2}Q_{k-1} - P_{k-1}Q_{k-2})a_k = (-1)^{k-1}a_k. \qquad \square$$

Предложение 1.25. Подходящие дроби с четными индексами образуют строго возрастающую последовательность, а подходящие дроби с нечетными индексами — строго убывающую последовательность.

Доказательство. Из предложения 1.24 следует, что:

$$\frac{P_k}{Q_k} - \frac{P_{k-2}}{Q_{k-2}} = \frac{(-1)^k a_k}{Q_kQ_{k-2}}.$$

Поэтому $\dfrac{P_k}{Q_k} > \dfrac{P_{k-2}}{Q_{k-2}}$ при четном k и $\dfrac{P_k}{Q_k} < \dfrac{P_{k-2}}{Q_{k-2}}$ при нечетном k. $\square$

Две подходящие дроби $\dfrac{P_{k-1}}{Q_{k-1}}$ и $\dfrac{P_k}{Q_k}$, у которых номер отличается на единицу, будем называть *соседними.*

Предложение 1.26. Из двух соседних подходящих дробей дробь с четным индексом всегда меньше дроби с нечетным индексом.

Доказательство. Учитывая предложение 1.21, имеем:

$$\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k} = \frac{(-1)^k}{Q_{k-1}Q_k}.$$

Поэтому $\dfrac{P_{k-1}}{Q_{k-1}} > \dfrac{P_k}{Q_k}$ при четном k и $\dfrac{P_{k-1}}{Q_{k-1}} < \dfrac{P_k}{Q_k}$ при нечетном k. $\square$

Предложение 1.27. Любая подходящая дробь с четным индексом меньше любой подходящей дроби с нечетным индексом.

Доказательство следует из предложений 1.25 и 1.26. □

Предложение 1.28. Расстояния (модули разностей) между соседними подходящими дробями строго уменьшаются с увеличением их номера.

Доказательство. Учитывая предложения 1.21 и 1.22, имеем:

$$\left|\frac{P_k}{Q_k} - \frac{P_{k+1}}{Q_{k+1}}\right| = \frac{1}{Q_k Q_{k+1}} < \frac{1}{Q_{k-1} Q_k} = \left|\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k}\right|. \qquad \square$$

Данные утверждения показывают, что подходящие дроби с четными и нечетными номерами являются левыми и правыми концами вложенных друг в друга отрезков:

$$\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \ldots < \frac{P_5}{Q_5} < \frac{P_3}{Q_3} < \frac{P_1}{Q_1},$$

причем последняя подходящая дробь $\dfrac{P_n}{Q_n} = \dfrac{a}{b}$ совпадает с величиной всей цепной дроби.

Предложение 1.29. При разложении a/b в цепную дробь:

1) a/b не меньше любой подходящей дроби с четным индексом;

2) a/b не больше любой подходящей дроби с нечетным индексом.

Доказательство следует из предложения 1.27. □

Следующее утверждение играет важную роль в вопросах приближенного представления чисел.

Предложение 1.30. При разложении a/b в цепную дробь:

$$\left|\frac{a}{b} - \frac{P_k}{Q_k}\right| < \frac{1}{Q_k Q_{k+1}}, \quad k = 0, \ldots, n-1.$$

Доказательство следует из предложения 1.28, так как a/b принадлежит отрезку $\left[\dfrac{P_k}{Q_k}, \dfrac{P_{k+1}}{Q_{k+1}}\right]$ при четном k или отрезку $\left[\dfrac{P_{k+1}}{Q_{k+1}}, \dfrac{P_k}{Q_k}\right]$ при нечетном k. □

1.8.2. Бесконечные цепные дроби

Определение 1.10. *Бесконечной цепной дробью* называется выражение вида:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}}, \tag{1.22}$$

где $a_0 \in \mathbb{Z}$, $a_k \in \mathbb{N}$, $k \geqslant 1$.

Для краткости записи выражение (1.22) будем иногда записывать в таком виде:

$$[a_0; a_1, a_2, \ldots].$$

Определение 1.11. *k-ой подходящей дробью* к бесконечной цепной дроби (1.22) будем называть конечную цепную дробь:

$$A_k = [a_0; a_1, \ldots, a_k].$$

Рассмотрим две последовательности чисел:

$$\{P_k\}_{k \geqslant -1}, \quad \{Q_k\}_{k \geqslant -1},$$

которые определим такими рекуррентными соотношениями при $k \geqslant 1$:

$$\begin{aligned} P_k &= P_{k-2} + P_{k-1} a_k, \\ Q_k &= Q_{k-2} + Q_{k-1} a_k \end{aligned} \tag{1.23}$$

и начальными условиями:

$$\begin{aligned} P_{-1} &= 1, \quad P_0 = a_0, \\ Q_{-1} &= 0, \quad Q_0 = 1. \end{aligned} \tag{1.24}$$

Предложение 1.31. Пусть a_0, a_1,... — элементы цепной дроби (1.22). Тогда для любого $k \geqslant 0$ выполнено равенство $A_k = \dfrac{P_k}{Q_k}$, где величины P_k и Q_k определяются рекуррентными соотношениями (1.23) с начальными условиями (1.24).

Доказательство следует из предложения 1.18.

Определение 1.12. Бесконечная цепная дробь (1.22) называется сходящейся, если существует предел ее подходящих дробей, т.е.:

$$\lim_{k\to\infty} A_k.$$

Величиной бесконечной сходящейся цепной дроби (1.22) называется предел ее подходящих дробей, т.е. такое число α, что $\lim_{k\to\infty} A_k = \alpha$.

Если величина (1.22) равна α, то будем записывать это в виде:

$$\alpha = [a_0; a_1, a_2, \ldots].$$

Свойства подходящих дробей, сформулированных в предложениях 1.18-1.28, справедливы и для бесконечных цепных дробей. Действительно, для любого натурального n подходящие дроби A_0, A_1, …, A_n к бесконечной дроби (1.22) являются также подходящими дробями к конечной цепной дроби:

$$[a_0; a_1, \ldots, a_{n+1}].$$

Сформулируем наиболее существенные свойства.

Предложение 1.32. Для любого $k \geqslant 0$ верно следующее равенство:

$$P_{k-1}Q_k - P_kQ_{k-1} = (-1)^k. \tag{1.25}$$

Предложение 1.33. Для любого $k \geqslant 1$ выполнены следующие равенства:

$$\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k} = \frac{(-1)^k}{Q_{k-1}Q_k},$$
$$\left|\frac{P_{k-1}}{Q_{k-1}} - \frac{P_k}{Q_k}\right| = \frac{1}{Q_{k-1}Q_k}.$$

Предложение 1.34. Последовательность $\{Q_k\}_{k\geqslant -1}$, начиная с $k = 1$, монотонно неограниченно возрастает:

$$1 = Q_0 \leqslant Q_1 < Q_2 < \ldots.$$

Предложение 1.35. Последовательность $\{P_k\}_{k\geqslant -1}$ положительной бесконечной цепной дроби, начиная с $k = 0$, монотонно неограниченно возрастает:

$$P_0 < P_1 < P_2 < \ldots.$$

Предложение 1.36. Модули расстояний между соседними подходящими дробями монотонно уменьшается с увеличением номера и стремится к нулю.

Предложение 1.37. Подходящие дроби с четными и нечетными номерами образуют систему концов вложенных друг в друга отрезков:

$$\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \ldots < \frac{P_5}{Q_5} < \frac{P_3}{Q_3} < \frac{P_1}{Q_1}.$$

Теорема 1.22. Любая бесконечная цепная дробь сходится.

Доказательство следует из предложений 1.36, 1.37 и известной леммы математического анализа (принцип вложенных отрезков Коши-Кантора). □

Замечание 1.4. Из сказанного выше следует, что величина бесконечной цепной дроби больше любой подходящей дроби с четным индексом и меньше любой подходящей дроби с нечетным индексом:

$$\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \ldots < \alpha < \ldots < \frac{P_5}{Q_5} < \frac{P_3}{Q_3} < \frac{P_1}{Q_1}.$$

Определение 1.13. Пусть $\alpha = [a_0; a_1, a_2, \ldots]$. *Полными частными* в разложении α будем называть величины b_0, b_1, b_2, $\ldots$, определенные равенствами:

$$\alpha = [a_0; a_1, \ldots, a_k, b_{k+1}], \quad k \geqslant 0,$$

$\alpha = b_0$ при $k = -1$.

Предложение 1.38. Пусть $\alpha = [a_0; a_1, a_2, \ldots]$, b_{k+1} — полное частное в разложении α. Тогда для любого $k \geqslant 0$ выполнены равенства:

$$\alpha = \frac{P_{k-1} + P_k b_{k+1}}{Q_{k-1} + Q_k b_{k+1}}, \tag{1.26}$$

$$b_{k+1} = \frac{P_{k-1} - \alpha Q_{k-1}}{\alpha Q_k - P_k}. \tag{1.27}$$

Доказательство. Из равенств:

$$A_{k+1} = \frac{P_{k+1}}{Q_{k+1}} = [a_0; a_1, \ldots, a_k, a_{k+1}],$$
$$\alpha = [a_0; a_1, \ldots, a_k, b_{k+1}],$$

видно, что если в A_{k+1} заменить a_{k+1} на b_{k+1}, то получим α. При этом:

$$\frac{P_{k+1}}{Q_{k+1}} = \frac{P_{k-1} + P_k a_{k+1}}{Q_{k-1} + Q_k a_{k+1}},$$

где P_{k-1}, P_k, Q_{k-1}, Q_k не зависят от a_{k+1}. Поэтому:

$$\alpha = \frac{P_{k+1}}{Q_{k+1}}\bigg|_{a_{k+1} \to b_{k+1}} = \frac{P_{k-1} + P_k b_{k+1}}{Q_{k-1} + Q_k b_{k+1}}.$$

Равенство (1.27) следует из (1.26). □

Определение 1.14. *Разложением действительного числа α в цепную дробь* называется представление α в виде

$$\alpha = [a_0; a_1, a_2, \ldots],$$

где a_0, a_1, a_2, $\ldots$ — конечная или бесконечная последовательность целых чисел, такая, что при $k \geqslant 1$ все $a_k \geqslant 1$, а в случае конечного разложения последний элемент $a_n > 1$.

Предложение 1.39. Пусть разложение α в цепную дробь имеет вид $\alpha = [a_0; a_1, a_2, \ldots]$. Обозначим $c_k = [a_k; a_{k+1}, \ldots]$. Тогда:

1. $\alpha = [a_0; a_1, \ldots, a_k, c_{k+1}]$, т.е. $c_k = b_k$ представляет собой k-е полное частное в разложении α.

2. $a_k = [c_k]$ для любого k, где $[\,]$ — целая часть числа.

Доказательство. 1. Для конечной цепной дроби это равенство очевидно. Пусть цепная дробь бесконечна. Из предложения 1.37 следует, что $c_k > 1$ для любого $k \geqslant 1$. Равенство $c_k = b_k$ следует из известных свойств предела (предел суммы, предел частного).

2. Если цепная дробь конечна и a_n — ее последний элемент, то $a_n = b_n = [b_n]$. Если a_k не является последним элементом, то:

$$b_{k+1} = c_{k+1} = [a_{k+1}; a_{k+2}, \ldots] > 1, \quad 0 < \frac{1}{c_{k+1}} < 1.$$

Так как $c_k = a_k + \dfrac{1}{c_{k+1}}$, то $a_k = [c_k]$. $\square$

Пример 1.5. 1. Найти величину цепной дроби

$$\alpha = [1; 3, 1, 3, \ldots],$$

где все дальнейшие элементы равны последовательно 1 и 3. Учитывая предложение 1.39, получаем:

$$\alpha = [1; 3, \alpha] = \frac{4\alpha + 1}{3\alpha + 1},$$

$$3\alpha^2 - 3\alpha - 1 = 0, \quad \alpha = \frac{3 \pm \sqrt{21}}{6}.$$

Так как $\alpha > 0$, то $\alpha = \dfrac{3 + \sqrt{21}}{6}$.

2. Найти величину цепной дроби

$$\alpha = [1; 2, 3, 4, 1, 2, 3, 4, \ldots],$$

где все дальнейшие элементы равны последовательно 1, 2, 3, 4.

Учитывая предложения 1.38 и 1.39, получаем:

$$\alpha = [1; 2, 3, 4, \alpha], \quad \alpha = \frac{P_3\alpha + P_2}{Q_3\alpha + Q_2}.$$

Составим таблицу значений P_k и Q_k при $k = -1, 0, 1, 2, 3$:

k	-1	0	1	2	3
a_k	-	1	2	3	4
P_k	1	1	3	10	43
Q_k	0	1	2	7	30

Поэтому:

$$\alpha = \frac{43\alpha + 10}{30\alpha + 7}, \quad 30\alpha^2 - 36\alpha - 10 = 0, \quad \alpha = \frac{9 + 2\sqrt{39}}{15}.$$

3. Найти величину цепной дроби

$$\alpha = [2; 1, 3, 1, 2, 1, 1, 2, 1, 1, \ldots],$$

где все дальнейшие элементы, начиная с a_4, равны последовательно 2, 1, 1.

В этом случае:

$$b_4 = [2; 1, 1, 2, 1, 1, \ldots] = [2; 1, 1, b_4],$$

Для нахождения b_4 используем прием предыдущего примера. Получим $b_4 = \dfrac{2 + \sqrt{10}}{2}$.

Тогда:

$$\alpha = [2; 1, 3, 1, b_4] = \frac{P_2 + P_3 b_4}{Q_2 + Q_3 b_4} = \frac{100 + \sqrt{10}}{37}.$$

Пример 1.6. Существует интересное число (известное ещё издревле), у которого все a_k равны 1:

$$\alpha = [1; 1, 1, \ldots].$$

Получаем:

$$\alpha = 1 + \frac{1}{\alpha}, \quad \alpha^2 - \alpha - 1 = 0, \quad \alpha = \frac{1 + \sqrt{5}}{2} \approx 1.6.$$

Это число имеет собственное имя — *золотое сечение*. Например, открытки делают в форме прямоугольника, отношение сторон которого равно этому числу. Если от такого прямоугольника отрезать квадрат со стороной, равной меньшей стороне прямоугольника, то оставшийся прямоугольник подобен исходному. Это и есть условие того, что отношение сторон равно золотому сечению. Если снова отрезать квадратик, снова получится прямоугольник, подобный исходному и т.д.

Теорема 1.23. Любое действительное число равно некоторой цепной дроби. Причем для каждого такого числа существует одна и только одна цепная дробь, равная данному числу.

Доказательство. Пусть $\alpha \in \mathbb{R}$. Если α — рациональное число, то справедливость утверждения следует из теоремы 1.21. Пусть α — иррациональное число.

Обозначим $a_0 = [\alpha]$ — целая часть числа, $b_1 = \dfrac{1}{\alpha - a_0}$. Получим $\alpha = a_0 + \dfrac{1}{b_1}$, где b_1 — иррациональное число, причем $b_1 > 1$.

Отсюда видно, что для любого иррационального числа α существует такое целое число $a_0 = [\alpha]$ и иррациональное $b_1 > 1$, что $\alpha = a_0 + \dfrac{1}{b_1}$. Найдем таким же образом для b_1 целое число $a_1 = [b_1] \geqslant 1$ и иррациональное $b_2 > 1$ и т.д.:

$$\begin{aligned}
\alpha &= a_0 + \frac{1}{b_1}, \quad & a_0 &= [\alpha], \\
b_1 &= a_1 + \frac{1}{b_2}, \quad & a_1 &= [b_1], \\
&\dots & &\dots \\
b_k &= a_k + \frac{1}{b_{k+1}}, & a_k &= [b_k], \\
&\dots & &\dots,
\end{aligned}$$

где иррациональные числа $b_k > 1$ для любого $k \geqslant 1$, поэтому $a_k = [b_k] \geqslant 1$, $k \geqslant 1$.

Числа a_0, a_1, a_2, ... образуют бесконечную последовательность целых чисел, причем $a_k \geqslant 1$ для всех $k \geqslant 1$. Поэтому эти числа можно взять в качестве элементов бесконечной цепной дроби $[a_0; a_1, a_2, \dots]$, которая сходится (теорема 1.22).

Покажем, что величина этой бесконечной цепной дроби равна исходному числу α. Действительно, так как:

$$\alpha = [a_0; a_1, \dots, a_k, b_{k+1}],$$

$$\alpha = \frac{P_{k-1} + P_k b_{k+1}}{Q_{k-1} + Q_k b_{k+1}},$$

$b_{k+1} > 1$, то:

$$\left|\alpha - \frac{P_k}{Q_k}\right| = \left|\frac{P_{k-1} + P_k b_{k+1}}{Q_{k-1} + Q_k b_{k+1}} - \frac{P_k}{Q_k}\right| =$$

$$= \left|\frac{P_{k-1}Q_k - P_k Q_{k-1}}{(Q_k b_{k+1} + Q_{k-1})Q_k}\right| = \frac{1}{(Q_k b_{k+1} + Q_{k-1})Q_k} < \frac{1}{Q_k^2}.$$

В данных равенствах использовались предложения 1.32 и 1.38. Учитывая предложение 1.34, получаем, что $\lim_{k\to\infty} \dfrac{P_k}{Q_k} = \alpha$.

Покажем единственность такого разложения. Если цепная дробь конечная, то единственность следует из теоремы 1.21. Пусть:

$$\alpha = [a_0; a_1, \ldots, a_k, \ldots] = [\widetilde{a}_0; \widetilde{a}_1, \ldots, \widetilde{a}_k, \ldots],$$

где хотя бы одна цепная дробь бесконечная.

Предположим, что эти две цепные дроби отличаются хотя бы одним элементом. Обозначим через k минимальный индекс, при котором $a_k \neq \widetilde{a}_k$:

$$a_0 = \widetilde{a}_0, \ldots, \ a_{k-1} = \widetilde{a}_{k-1}, \ a_k \neq \widetilde{a}_k.$$

Обозначим $c_k = [a_k; a_{k+1}, \ldots]$, $\widetilde{c}_k = [\widetilde{a}_k; \widetilde{a}_{k+1}, \ldots]$. Из равенства (предложение 1.39):

$$\alpha = [a_0; a_1, \ldots, a_{k-1}, c_k] = [\widetilde{a}_0; \widetilde{a}_1, \ldots, \widetilde{a}_{k-1}, \widetilde{c}_k]$$

следует, что $c_k = \widetilde{c}_k$. При этом:

$$a_k = [c_k] = [\widetilde{c}_k] = \widetilde{a}_k.$$

Противоречие. □

Теорема 1.23 дает алгоритм разложения действительных чисел в цепную дробь.

Пример 1.7. 1. Разложим в цепную дробь число $1 + \sqrt{2}$.

Вычисляем:

$$a_0 = \left[1 + \sqrt{2}\right] = 2, \quad b_1 = \frac{1}{1 + \sqrt{2} - 2} = 1 + \sqrt{2}.$$

Так как $\alpha = b_1$, то $a_k = 2$ для любого $k \geqslant 1$. Поэтому:

$$1 + \sqrt{2} = [2; 2, 2, \ldots].$$

2. Разложим в цепную дробь число $\alpha = \dfrac{4 + \sqrt{2}}{4}$. Вычисляем:

$$a_0 = [\alpha] = 1, \quad b_1 = \frac{1}{\alpha - a_0} = 2\sqrt{2},$$

$$a_1 = [b_1] = 2, \quad b_2 = \frac{1}{b_1 - a_1} = \frac{1 + \sqrt{2}}{2},$$

$$a_2 = [b_2] = 1, \quad b_3 = \frac{1}{b_2 - a_2} = 2 + 2\sqrt{2},$$

$$a_3 = [b_3] = 4, \quad b_4 = \frac{1}{b_3 - a_3} = \frac{1 + \sqrt{2}}{2}.$$

Так как $b_2 = b_4$, то:

$$b_2 = [1; 4, b_4] = [1; 4, b_2], \quad \alpha = [1; 2, b_2] = [1; 2, \underbrace{1, 4}, 1, 4, \ldots].$$

1.8.3. Квадратичные иррациональности

Определение 1.15. Бесконечная цепная дробь:

$$\alpha = [a_0; a_1, a_2, \ldots]$$

называется *периодической*, если существуют такие целые числа $N \geqslant 0$ и $t > 0$, что для любого $k \geqslant N$ выполнено равенство $a_{k+t} = a_k$:

$$\alpha = [a_0; a_1, \ldots, a_N, \ldots, a_{N+t-1}, a_N, \ldots, a_{N+t-1}, \ldots].$$

Определение 1.16. Действительное число α называется *квадратичной иррациональностью*, если найдутся такие взаимно простые целые числа $a > 0$, b, c, что значение $b^2 - 4ac > 0$ не является полным квадратом, а число α является одним из корней многочлена $f(x) = ax^2 + bx + c$, т.е. $f(\alpha) = 0$.

Из определения 1.16 следует, что любая квадратичная иррациональность может быть представлена в виде $\frac{A+\sqrt{D}}{B}$, где A, B, D — целые числа, $D = b^2 - 4ac$ не является полным квадратом и:

$$\begin{cases} A = -b, & B = 2a, \\ A = b, & B = -2b, \end{cases}$$

в зависимости от того, какой из двух корней многочлена $f(x)$ выбирается.

Теорема 1.24 (**Лагранж**). Иррациональное число α представляется бесконечной периодической цепной дробью тогда и только тогда, когда α является квадратичной иррациональностью.

Доказательство. Пусть иррациональное число α представляется бесконечной периодической цепной дробью. Так как из равенств $a_{k+t} = a_k$ следует $b_{k+t} = b_k$ для любого $k \geqslant N$, то, учитывая предложение 1.38, получаем:

$$\alpha = \frac{P_{k-2} + b_k P_{k-1}}{Q_{k-2} + b_k Q_{k-1}} = \frac{P_{k-2+t} + b_{k+t} P_{k-1+t}}{Q_{k-2+t} + b_{k+t} Q_{k-1+t}} = \frac{P_{k-2+t} + b_k P_{k-1+t}}{Q_{k-2+t} + b_k Q_{k-1+t}},$$

откуда:

$$\frac{P_{k-2} + b_k P_{k-1}}{Q_{k-2} + b_k Q_{k-1}} = \frac{P_{k-2+t} + b_k P_{k-1+t}}{Q_{k-2+t} + b_k Q_{k-1+t}}.$$

Поэтому b_k является корнем квадратного уравнения:

$$(P_{k-1}Q_{k-1+t} - Q_{k-1}P_{k-1+t})x^2 +$$

$$+(P_{k-2}Q_{k-1+t} + P_{k-1}Q_{k-2+t} - Q_{k-2}P_{k-1+t} - Q_{k-1}P_{k-2+t})x +$$

$$+(P_{k-2}Q_{k-2+t} - Q_{k-2}P_{k-2+t}) = 0$$

с целыми коэффициентами. Следовательно, b_k является квадратичной иррациональностью. А из равенства:

$$\alpha = \frac{P_{k-2} + b_k P_{k-1}}{Q_{k-2} + b_k Q_{k-1}}$$

следует, что α — квадратичная иррациональность.

Обратно, пусть α — иррациональный корень квадратного уравнения с целыми коэффициентами:

$$ax^2 + bx + c = 0 \tag{1.28}$$

и $\alpha = [a_0; a_1, a_2, \ldots]$ — его представление бесконечной цепной дробью. Запишем α в виде:

$$\alpha = \frac{P_{k-2} + b_k P_{k-1}}{Q_{k-2} + b_k Q_{k-1}}, \quad k \geqslant 2.$$

Подставив это выражение в формулу (1.28), получим, что b_k является корнем квадратного уравнения:

$$A_k x^2 + B_k x + C_k = 0, \tag{1.29}$$

в котором:

$$A_k = aP_{k-1}^2 + bP_{k-1}Q_{k-1} + cQ_{k-1}^2,$$

$$B_k = 2aP_{k-1}P_{k-2} + b(P_{k-1}Q_{k-2} + Q_{k-1}P_{k-2}) + 2cQ_{k-1}Q_{k-2},$$
$$C_k = aP_{k-2}^2 + bP_{k-2}Q_{k-2} + cQ_{k-2}^2.$$

Заметим, что $C_k = A_{k-1}$. Найдем дискриминант квадратного уравнения (1.29):

$$D_k = B_k^2 - 4A_kC_k = (b^2 - 4ac)(P_{k-1}Q_{k-2} - Q_{k-1}P_{k-2})^2 = b^2 - 4ac,$$

т.е. дискриминант уравнения (1.29) совпадает с дискриминантом уравнения (1.28) и не зависит от k. Из неравенств:

$$\left|\alpha - \frac{P_{k-1}}{Q_{k-1}}\right| < \left|\frac{P_k}{Q_k} - \frac{P_{k-1}}{Q_{k-1}}\right| = \frac{1}{Q_{k-1}Q_k} < \frac{1}{Q_{k-1}^2}$$

следует, что число $\alpha Q_{k-1} - P_{k-1}$ представляется в виде:

$$\alpha Q_{k-1} - P_{k-1} = \frac{\varepsilon_{k-1}}{Q_{k-1}}, \quad |\varepsilon_{k-1}| < 1.$$

Поэтому:

$$P_{k-1} = \alpha Q_{k-1} - \frac{\varepsilon_{k-1}}{Q_{k-1}}$$

и

$$A_k = aP_{k-1}^2 + bP_{k-1}Q_{k-1} + cQ_{k-1}^2 =$$
$$a\left(\alpha Q_{k-1} - \frac{\varepsilon_{k-1}}{Q_{k-1}}\right)^2 + b\left(\alpha Q_{k-1} - \frac{\varepsilon_{k-1}}{Q_{k-1}}\right)Q_{k-1} + cQ_{k-1}^2 =$$
$$= Q_{k-1}^2(a\alpha^2 + b\alpha + c) - 2a\alpha\varepsilon_{k-1} + a\frac{\varepsilon_{k-1}^2}{Q_{k-1}^2} - b\varepsilon_{k-1} =$$
$$= Q_{k-1}\cdot 0 - 2a\alpha\varepsilon_{k-1} + a\frac{\varepsilon_{k-1}^2}{Q_{k-1}^2} - b\varepsilon_{k-1} =$$
$$= -2a\alpha\varepsilon_{k-1} + a\frac{\varepsilon_{k-1}^2}{Q_{k-1}^2} - b\varepsilon_{k-1}.$$

Отсюда следует, что:

$$|A_k| < 2|a\alpha| + |a| + |b|, \quad |C_k| = |A_{k-1}| < 2|a\alpha| + |a| + |b|.$$

Таким образом, величины A_k и C_k ограничены и при различных значений k принимают лишь конечное число различных значений. А так как $B_k^2 - 4A_kC_k = b^2 - 4ac$, то и B_k принимает лишь конечное число различных значений. Поэтому при

$k \in \mathbb{N}$ существует лишь конечное число уравнений вида (1.29), корнями которых являются b_k. Таким образом, найдутся такие натуральные k и t, для которых $b_k = b_{k+t}$. Отсюда и из единственности представления иррационального числа бесконечной цепной дробью следует периодичность данной дроби. $\square$

1.9. Мультипликативные функции

Определение 1.17. Функция $\theta : \mathbb{N} \to \mathbb{C}$ называется мультипликативной, если она удовлетворяет следующим условиям:

1. Данная функция определена на множестве всех натуральных чисел, причем не является тождественно нулевой, т.е. хотя бы для одного натурального числа функция θ не равна нулю.

2. Для любых натуральных взаимно простых чисел a и b выполнено:

$$\theta(ab) = \theta(a)\theta(b).$$

Примеры мультипликативных функций.

1. $\theta(a) = 1$ для любого натурального a.
2. $\theta(a) = a$ для любого натурального a.
3. $\theta(a) = \dfrac{1}{a}$ для любого натурального a.

Заметим, что если θ — некоторая мультипликативная функция, то $\theta(1) = 1$. Действительно, пусть a_0 — такое натуральное число, что $\theta(a_0) \neq 0$. Тогда:

$$\theta(a_0) = \theta(a_0 \cdot 1) = \theta(a_0)\theta(1).$$

Поэтому $\theta(1) = 1$. Заметим также, что из предложения 1.6 следует, что если a_1, a_2,..., a_n — некоторые попарно взаимно простые числа, то:

$$\theta(a_1 a_2 \dots a_n) = \theta(a_1)\theta(a_2)\dots\theta(a_n).$$

Используя свойство коммутативности умножения целых чисел, легко показать, что произведение двух мультипликативных функций, определенное по правилу:

$$(\theta_1\theta_2)(a) = \theta_1(a)\theta_2(a),$$

также будет являться мультипликативной функцией.

Лемма 1.1. Пусть θ — некоторая мультипликативная функция и $a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ — каноническое разложение натурального числа a. Тогда сумма по всем делителям d числа a представима в следующем виде:

$$\sum_{d \mid a} \theta(d) = (1 + \theta(p_1) + \dots + \theta(p_1^{\alpha_1})) \dots (1 + \theta(p_n) + \dots + \theta(p_n^{\alpha_n})), \tag{1.30}$$

причем если $a = 1$, то правая часть равна 1.

Доказательство. Из теоремы 1.17 следует, что все делители числа a имеют такой вид:

$$d = p_1^{\beta_1} p_2^{\beta_2} \dots p_n^{\beta_n}, \quad \text{где}$$

$$0 \leqslant \beta_1 \leqslant \alpha_1, \ 0 \leqslant \beta_2 \leqslant \alpha_2, \dots, \ 0 \leqslant \beta_n \leqslant \alpha_n.$$

Поэтому:

$$\sum_{d \mid a} \theta(d) = \sum_{\beta_1=0}^{\alpha_1} \dots \sum_{\beta_n=0}^{\alpha_n} \theta(p_1^{\beta_1} p_2^{\beta_2} \dots p_n^{\beta_n}) =$$

$$= \sum_{\beta_1=0}^{\alpha_1} \dots \sum_{\beta_n=0}^{\alpha_n} \theta(p_1^{\beta_1}) \theta(p_2^{\beta_2}) \dots \theta(p_n^{\beta_n}) =$$

$$= \left(\sum_{\beta_1=0}^{\alpha_1} \theta(p_1^{\beta_1}) \right) \dots \left(\sum_{\beta_n=0}^{\alpha_n} \theta(p_n^{\beta_n}) \right). \qquad \square$$

Число делителей и сумма делителей. Если рассмотреть случай $\theta(a) = 1$, тогда тождество (1.30) будет иметь такой вид:

$$\tau(a) = (1 + \alpha_1)(1 + \alpha_2) \dots (1 + \alpha_n),$$

где $\tau(a)$ — количество делителей числа a.

Если же рассмотреть случай $\theta(a) = a$, тогда тождество (1.30) будет иметь следующий вид:

$$S(a) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \frac{p_2^{\alpha_2+1} - 1}{p_2 - 1} \dots \frac{p_n^{\alpha_n+1} - 1}{p_n - 1},$$

где $S(a)$ — сумма делителей числа a.

1.9.1. Функция Мебиуса

Определение 1.18. *Функция Мебиуса* — мультипликативная функция, которая определяется следующими равенствами:

$$\mu(1) = 1, \quad \mu(p) = -1,$$
$$\mu(p^\alpha) = 0, \quad \text{если } \alpha > 1,$$
$$\mu(p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}) = \mu(p_1^{\alpha_1})\mu(p_1^{\alpha_2}) \dots \mu(p_n^{\alpha_n}).$$

Поэтому если каноническое разложение $a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ содержит некоторое значение $\alpha_i > 1$, то $\mu(a) = 0$.

Из тождества (1.30) непосредственно следует лемма 1.2.

Лемма 1.2. Пусть $\theta(a)$ — некоторая мультипликативная функция и $a = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ — каноническое разложение числа a. Тогда:

$$\sum_{d \mid a} \mu(d)\theta(d) = (1 - \theta(p_1)) \dots (1 - \theta(p_n)),$$

причем правая часть равна 1, если $a = 1$. В частности:

1. Если $\theta(a) = 1$, то:

$$\sum_{d \mid a} \mu(d) = \begin{cases} 0, & a > 1, \\ 1, & a = 1. \end{cases}$$

2. Если $\theta(a) = \dfrac{1}{a}$, то:

$$\sum_{d \mid a} \frac{\mu(d)}{d} = \begin{cases} \left(1 - \frac{1}{p_1}\right) \dots \left(1 - \frac{1}{p_n}\right), & \text{если } a > 1, \\ 1, & \text{если } a = 1. \end{cases}$$

Лемма 1.3. Пусть некоторому набору натуральных чисел x_1, $x_2, \dots, x_n$ поставлены в соответствие (в общем случае комплексные) числа f_1, $f_2, \dots, f_n$:

$$x_1 \longleftrightarrow f_1, \quad x_2 \longleftrightarrow f_2, \dots, \ x_n \longleftrightarrow f_n.$$

Пусть $\widetilde{S}$ — сумма тех значений f_i, которые отвечают значениям x_i, равным 1, а S_d — сумма тех значений f_i, которые отвечают значениям x_i, кратным d. Тогда:

$$\widetilde{S} = \sum_d \mu(d) S_d,$$

где d пробегает все натуральные числа от 1 до $\max\{x_1, \ldots, x_n\}$ и делит хотя бы одно из чисел x_1, x_2,..., x_n.

Доказательство. Из леммы 1.2 (пункт 1) следует такое равенство:

$$\widetilde{S} = f_1 \sum_{d \mid x_1} \mu(d) + \ldots + f_n \sum_{d \mid x_n} \mu(d).$$

Если в правой части данного равенства собрать вместе слагаемые с одинаковыми значениями d, при этом вынося $\mu(d)$ за скобку, получим сумму только тех значений f_i, которые отвечают значениям x_i, кратным d, т.е. получим сумму S_d. □

1.9.2. Функция Эйлера

Определение 1.19. *Функция Эйлера* $\varphi(a)$ — функция, определенная на множестве всех натуральных чисел, значение которой от аргумента a представляет собою количество чисел ряда:

$$0,\ 1,\ 2, \ldots,\ a-1,$$

взаимно простых с числом a.

Например, $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$, $\varphi(5) = 4$.

Теорема 1.25. Пусть $a = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$ — каноническое разложение числа a. Тогда:

$$\varphi(a) = a\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\ldots\left(1 - \frac{1}{p_n}\right). \qquad (1.31)$$

Из данного равенства следует также, что:

$$\varphi(a) = (p_1^{\alpha_1} - p_1^{\alpha_1 - 1})(p_2^{\alpha_2} - p_2^{\alpha_2 - 1}) \ldots (p_n^{\alpha_n} - p_n^{\alpha_n - 1}). \qquad (1.32)$$

Доказательство. Воспользуемся леммой 1.3. Пусть:

$$x_0 = (0, a),\ x_1 = (1, a), \ldots,\ x_{a-1} = (a-1, a).$$

Каждому из данных значений поставим в соответствие единицу:

$$x_0 \longleftrightarrow 1,\quad x_1 \longleftrightarrow 1, \ldots,\ x_{a-1} \longleftrightarrow 1,$$

т.е. $f_0 = f_1 = \ldots = f_{a-1} = 1$. Тогда число $\widetilde{S}$ будет означать количество таких элементов x_i, которые равны 1, т.е. $(i, a) = 1$.

Поэтому в данном случае $\varphi(a) = \widetilde{S}$. Число же S_d будет означать количество таких x_i, которые кратны d. Но если $x_i = (i, a)$ делится на d, то, в частности, a делится на d. Поэтому все элементы из набора x_0, x_1,..., x_{a-1}, делящиеся на d, можно записать в таком виде:

$$(0d, a),\ (d, a),\ (2d, a), \ldots,\ \left(\left(\frac{a}{d} - 1\right) d, a\right).$$

Следовательно, $S_d = \dfrac{a}{d}$ и имеет место такое равенство:

$$\varphi(a) = \sum_{d \mid a} \mu(d)\frac{a}{d} = a \sum_{d \mid a} \frac{\mu(d)}{d}.$$

Из данного равенства и леммы 1.2 (пункт 2) следует формула (1.31), из которой следует формула (1.32). □

Следствие 1.7. Функция Эйлера является мультипликативной, и для любого натурального a выполнено равенство:

$$\sum_{d \mid a} \varphi(d) = a,$$

где суммирование берется по всем делителям числа a.

Доказательство. Мультипликативность функции Эйлера следует из теоремы 1.25. Покажем вторую часть утверждения.

Пусть $a = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$ — каноническое разложение числа a. Из леммы 1.1 следует, что:

$$\sum_{d \mid a} \varphi(d) = (1 + \varphi(p_1) + \ldots + \varphi(p_1^{\alpha_1})) \ldots (1 + \varphi(p_n) + \ldots + \varphi(p_n^{\alpha_n})).$$

Также из теоремы 1.25 следует, что для любого простого числа p и любого натурального n выполнено равенство:

$$\varphi(p^n) = p^n - p^{n-1}.$$

Поэтому:

$$\sum_{i=0}^{n} \varphi(p^i) = 1 + (p - 1) + (p^2 - p) + \ldots + (p^n - p^{n-1}) = p^n.$$

Следовательно:

$$\sum_{d \mid a} \varphi(d) = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n} = a.$$ □

1.10. Сравнения

Пусть m — некоторое натуральное число.

Определение 1.20. Целые числа a и b называются *сравнимыми по модулю* m, если разность $a - b$ делится на m.

В этом случае будем записывать:

$$a \equiv b \pmod{m}.$$

Данное соотношение читается «a сравнимо с b по модулю m».

Лемма 1.4. Пусть a и b — некоторые целые числа и m — некоторое натуральное число. Тогда следующие условия эквивалентны.

(i) a сравнимо с b по модулю m.

(ii) a и b имеют одинаковые остатки при делении на m.

(iii) имеет место представление $a = b + mt$, где t — некоторое целое число.

Доказательство. Пусть $a \equiv b \pmod{m}$. Представим a и b в виде:

$$a = mq_1 + r_1, \quad b = mq_2 + r_2, \quad \text{где} \quad 0 \leqslant r_1, r_2 < m.$$

Без ограничения общности можно считать, что $r_1 \geqslant r_2$. Тогда:

$$a - b = m(q_1 - q_2) + (r_1 - r_2), \quad \text{причем } 0 \leqslant r_1 - r_2 < m.$$

По условию a сравнимо с b по модулю m, поэтому $m|(a - b)$. Также из теоремы 1.1 следует, что $r_1 - r_2 = 0$. Таким образом, из условия (i) следует (ii).

Пусть выполнено условие (ii) :

$$a = mq_1 + r, \quad b = mq_2 + r, \quad \text{где} \quad 0 \leqslant r < m.$$

Тогда $a - b = m(q_1 - q_2)$, что означает $a = b + m(q_1 - q_2)$. Поэтому условие (ii) влечет условие (iii).

Из условия (iii) очевидным образом следует условие (i). $\square$

Согласно данной лемме, в качестве определения сравнения можно дать такое эквивалентное определение.

Определение 1.21. Целые числа a и b называются *сравнимыми по модулю* m, если остатки от деления этих чисел на m равны.

Приведем некоторые свойства сравнений.

Теорема 1.26. 1. $a \equiv a \pmod m$ (рефлексивность отношения сравнимости).

2. Если $a \equiv b \pmod m$, то $b \equiv a \pmod m$ (симметричность отношения сравнимости).

3. Если $a \equiv b \pmod m$, $b \equiv c \pmod m$, то $a \equiv c \pmod m$ (транзитивность отношения сравнимости).

4. Если $a \equiv b \pmod m$ и k — произвольное целое число, то $ka \equiv kb \pmod m$.

5. Если $ka \equiv kb \pmod m$ и $(k, m) = 1$, то $a \equiv b \pmod m$.

6. Если $a \equiv b \pmod m$ и k — произвольное натуральное число, то $ka \equiv kb \pmod{km}$.

7. Если $ka \equiv kb \pmod{km}$ и k — произвольное натуральное число, то $a \equiv b \pmod m$.

8. Если $a \equiv b \pmod m$ и $c \equiv d \pmod m$, то $a+c \equiv b+d \pmod m$ и $a - c \equiv b - d \pmod m$.

8′. Если $a_1 \equiv b_1 \pmod m$, $a_2 \equiv b_2 \pmod m$,..., $a_n \equiv b_n \pmod m$, то $a_1+a_2+\ldots+a_n \equiv b_1+b_2+\ldots+b_n \pmod m$.

9. Если $a \equiv b \pmod m$ и $c \equiv d \pmod m$, то $ac \equiv bd \pmod m$.

9′. Если $a_1 \equiv b_1 \pmod m$, $a_2 \equiv b_2 \pmod m$,..., $a_n \equiv b_n \pmod m$, то $a_1a_2 \ldots a_n \equiv b_1b_2 \ldots b_n \pmod m$.

9″. Если $a \equiv b \pmod m$ и n — произвольное натуральное число, то $a^n \equiv b^n \pmod m$.

10. Если $a \equiv b \pmod m$ и $f(x) = c_0 + c_1x + \ldots + c_nx^n$ — произвольный многочлен над кольцом целых чисел ($f(x) \in \mathbb{Z}[x]$), то $f(a) \equiv f(b) \pmod m$.

11. Если $a_1 \equiv b_1 \pmod m$, $a_2 \equiv b_2 \pmod m$,..., $a_n \equiv b_n \pmod m$ и $f(x_1, x_2, \ldots, x_n)$ — произвольный многочлен от неизвестных x_1, x_2,..., x_n с целыми коэффициентами, то:

$$f(a_1, a_2, \ldots, a_n) \equiv f(b_1, b_2, \ldots, b_n) \pmod m.$$

12. Если $a \equiv b \pmod m$ и $d|m$, то $a \equiv b \pmod d$.

13. Если $a \equiv b \pmod{m_1}, \ldots, a \equiv b \pmod{m_k}$, то:

$$a \equiv b \pmod{[m_1, \ldots, m_k]}.$$

14. Если $a \equiv b \pmod{m}$, то $(a, m) = (b, m)$.

Пусть m — некоторое фиксированное натуральное число. Поскольку отношение сравнимости является отношением эквивалентности (теорема 1.26), то оно индуцирует разбиение множества целых чисел на классы эквивалентных элементов. При этом два целых числа a и b принадлежат одному и тому же классу тогда и только тогда, когда $a \equiv b \pmod{m}$. Поэтому одному и тому же классу будут принадлежать все целые числа, которые при делении на m дают в остатке одно и то же число r. Множество целых чисел можно представить в следующем виде:

$$\mathbb{Z} = Z_0 \cup Z_1 \cup \ldots \cup Z_{m-1},$$

где $Z_i = \{mq + i \mid q \in \mathbb{Z}\}$ и $Z_i \cap Z_j = \varnothing$, если $i \neq j$.

Определение 1.22. *Вычетом класса* называется любое число, принадлежащее данному классу.

Определение 1.23. *Полной системой вычетов* по некоторому модулю m называется совокупность чисел, взятых по одному из каждого класса.

Лемма 1.5. Любые m чисел, попарно несравнимых по модулю m, образуют полную систему вычетов по данному модулю.

Доказательство. Так как количество классов, сравнимых по модулю m элементов, ровно m и все числа, указанные в условии леммы, принадлежат к разным классам, то данные числа будут образовывать полную систему вычетов по модулю m. □

Лемма 1.6. Пусть a и b — некоторые целые числа, причем $(a, m) = 1$, где m — некоторое натуральное число. Пусть также x_1, $x_2, \ldots$, x_m — полная система вычетов по модулю m. Тогда числа:

$$ax_1 + b,\ ax_2 + b, \ldots,\ ax_m + b \tag{1.33}$$

будут также образовывать полную систему вычетов по модулю m.

Доказательство. Покажем, что любые два числа из совокупности (1.33) несравнимы по модулю m. Предположим противное. Пусть:

$$ax_i + b \equiv ax_j + b \pmod{m}$$

для некоторых i и j, причем $i \neq j$. Но тогда из теоремы 1.26 следует, что $ax_i \equiv ax_j \pmod{m}$. Так как $(a, m) = 1$, то по той же теореме получаем, что $x_i \equiv x_j \pmod{m}$. Противоречие. Поэтому числа из совокупности (1.33) попарно несравнимы по модулю m и их количество ровно m. Следовательно, по лемме 1.5 они образуют полную систему вычетов по модулю m. □

Из теоремы 1.26 видно, что числа одного и того же класса по модулю m имеют с числом m один и тот же наибольший общий делитель. Очень важны классы, для которых НОД равен 1, т.е. классы, содержащие числа, взаимно простые с модулем m.

Определение 1.24. *Приведенной системой вычетов* по некоторому модулю m называется совокупность всех таких чисел из полной системы вычетов, которые взаимно просты с модулем m.

Если в качестве полной системы вычетов рассмотреть совокупность чисел $0, 1, \ldots, m-1$, то число $\varphi(m)$ будет равно количеству чисел из приведенной системы вычетов.

Лемма 1.7. Любые $\varphi(m)$ чисел, попарно несравнимых по модулю m и взаимно простых с m, образуют приведенную систему вычетов по данному модулю.

Лемма 1.8. Пусть a — некоторое целое число с тем условием, что $(a, m) = 1$, где m — некоторое натуральное число. Пусть также $x_1, x_2, \ldots, x_{\varphi(m)}$ — приведенная система вычетов по модулю m. Тогда числа:

$$ax_1,\ ax_2, \ldots,\ ax_{\varphi(m)} \tag{1.34}$$

будут также образовывать приведенную систему вычетов по модулю m.

Доказательство. Количество чисел из совокупности (1.34) ровно $\varphi(m)$. Чтобы применить для данных чисел лемму 1.7, достаточно показать, что числа из совокупности (1.34) попарно несравнимы по модулю m и взаимно просты с m.

Действительно, данные числа являются попарно несравнимыми по модулю m (лемма 1.6 при $b = 0$). Также из предложения 1.5 следует, что $(ax_i, m) = (x_i, m) = 1$, $i = 1, 2, \ldots, \varphi(m)$. □

1.11. Теоремы Эйлера и Ферма

Теорема 1.27 (**Эйлер**). Для любого натурального m и любого целого a такого, что $(a, m) = 1$, справедливо сравнение:

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

Доказательство. Пусть x_1, x_2,..., $x_{\varphi(m)}$ — приведенная система вычетов по модулю m. Из леммы 1.8 следует, что числа ax_1, ax_2,..., $ax_{\varphi(m)}$ также образуют приведенную систему вычетов по модулю m. Установим взаимно однозначное соответствие между множествами чисел:

$$\{ax_1, ax_2, \ldots, ax_{\varphi(m)}\}, \quad \{x_1, x_2, \ldots, x_{\varphi(m)}\},$$

сопоставив каждому из чисел первого множества сравнимое с ним число по модулю m из второго множества:

$$ax_1 \equiv x_{i_1} \pmod{m},$$
$$ax_2 \equiv x_{i_2} \pmod{m},$$
$$\ldots$$
$$ax_{\varphi(m)} \equiv x_{i_{\varphi(m)}} \pmod{m},$$

где x_{i_1}, x_{i_2},..., $x_{i_{\varphi(m)}}$ — некоторым образом переставленные числа x_1, x_2,..., $x_{\varphi(m)}$. Перемножая все сравнения, получаем:

$$a^{\varphi(m)} x_1 x_2 \ldots x_{\varphi(m)} \equiv x_{i_1} x_{i_2} \ldots x_{i_{\varphi(m)}} \pmod{m}.$$

Так как $(x_i, m) = 1$ для всех $i = 1, 2, \ldots, \varphi(m)$, то из предложения 1.6 следует, что $(x_1 x_2 \ldots x_{\varphi(m)}, m) = 1$. Учитывая еще, что:

$$x_1 x_2 \ldots x_{\varphi(m)} = x_{i_1} x_{i_2} \ldots x_{i_{\varphi(m)}},$$

то из теоремы 1.26 следует сравнение $a^{\varphi(m)} \equiv 1 \pmod{m}$. $\square$

Из теоремы Эйлера непосредственно следует малая теорема Ферма.

Теорема 1.28 (Ферма). Для любого простого p и любого целого a, не делящегося на p, справедливо сравнение:

$$a^{p-1} \equiv 1 \pmod{p}.$$

Функция Эйлера $\varphi(m)$ не всегда является наименьшим положительным значением k, для которого $a^k \equiv 1 \pmod{m}$. Введем в рассмотрение обобщенную функцию Эйлера.

Определение 1.25. *Обобщенной функцией Эйлера* $L(m)$ называется функция, определенная для всех натуральных значений m следующим образом: $L(1) = 1$, а при $m > 1$:

$$L(m) = \left[p_1^{\alpha_1 - 1}(p_1 - 1), \ldots, p_n^{\alpha_n - 1}(p_n - 1)\right],$$

где $m = p_1^{\alpha_1} \ldots p_n^{\alpha_n}$ — каноническое разложение числа m.

Пример 1.8. $L(180) = L(2^2 \cdot 3^2 \cdot 5) = [2, 6, 4] = 12$. При этом $\varphi(180) = 2 \cdot 6 \cdot 4 = 48$.

При $m = p^\alpha$ значения $L(m)$ и $\varphi(m)$, очевидно, совпадают.

Теорема 1.29. Для любого натурального m и любого целого a такого, что $(a, m) = 1$, справедливо сравнение:

$$a^{L(m)} \equiv 1 \pmod{m}.$$

Доказательство. Пусть $m = p_1^{\alpha_1} \ldots p_n^{\alpha_n}$ — каноническое разложение числа m. По теореме Эйлера:

$$a^{p_i^{\alpha_i - 1}(p_i - 1)} \equiv 1 \pmod{p_i^{\alpha_i}}, \quad i = 1, \ldots, n.$$

Возведем обе части данного сравнения в степень $\dfrac{L(m)}{p_i^{\alpha_i - 1}(p_i - 1)}$:

$$a^{L(m)} \equiv 1 \pmod{p_i^{\alpha_i}}, \quad i = 1, \ldots, n.$$

Так как $[p_1^{\alpha_1}, \ldots, p_n^{\alpha_n}] = m$, то по теореме 1.26 $a^{L(m)} \equiv 1 \pmod{m}$. $\square$

1.12. Сравнения первой степени

Рассмотрим сравнение с одним неизвестным вида:

$$ax \equiv b \pmod{m}. \tag{1.35}$$

Если данное сравнение имеет решение при некотором x_0, то и все числа, лежащие в одном классе с x_0, будут решениями данного сравнения. Поэтому весь этот класс считается за одно решение. Исходя из данного соглашения, решить сравнение (1.35) — значит указать все значения из полной системы вычетов, которые удовлетворяют сравнению (1.35).

Теорема 1.30. Если $(a, m) = 1$, то сравнение (1.35) имеет, и притом единственное, решение.

Доказательство. Пусть x_0, $x_1, \ldots$, x_{m-1} — полная система вычетов. Из леммы 1.6 следует, что числа ax_0, $ax_1, \ldots$, ax_{m-1} также будут образовывать полную систему вычетов по модулю m. Значит, среди данных чисел будет одно и только одно число, которое лежит в одном классе с числом b. Обозначим это число через ax_k. Тогда $ax_k \equiv b \pmod{m}$. □

В следующей теореме приводятся способы нахождения решений сравнения (1.35) при $(a, m) = 1$.

Теорема 1.31. Пусть $(a, m) = 1$. Тогда представленные ниже классы будут являться решениями сравнения (1.35) :

(i) $x \equiv ba^{\varphi(m)-1} \pmod{m}$;

(ii) $x \equiv vb \pmod{m}$, где v — целое число из пары целых чисел u и v, для которых выполнено равенство $mu+av = 1$ (следствие 1.1);

(iii) $x \equiv (-1)^n bP_{n-1} \pmod{m}$, где $\dfrac{P_0}{Q_0}, \dfrac{P_1}{Q_1}, \ldots, \dfrac{P_n}{Q_n} = \dfrac{m}{a}$ — последовательность подходящих дробей разложения $\dfrac{m}{a}$ в конечную цепную дробь.

Доказательство. (i) Применяя теорему Эйлера, получим:

$$a(ba^{\varphi(m)-1}) = ba^{\varphi(m)} \equiv b \pmod{m}.$$

(*ii*) Так как $(a, m) = 1$, то найдутся такие целые u и v, для которых выполнено равенство $mu + av = 1$ (следствие 1.1). Из данного равенства следует, что $av \equiv 1 \pmod{m}$. Поэтому выполнено сравнение $avb \equiv b \pmod{m}$.

(*iii*) Так как $(a, m) = 1$ и $(P_n, Q_n) = 1$ (предложение 1.20), то из равенства двух несократимых дробей $\dfrac{m}{a} = \dfrac{P_n}{Q_n}$ следует, что $m = P_n$ и $a = Q_n$. Так как справедливо равенство (предложение 1.19):

$$P_{n-1}Q_n - P_nQ_{n-1} = (-1)^n,$$

которое можно записать в таком виде:

$$aP_{n-1} - mQ_{n-1} = (-1)^n,$$

то $aP_{n-1} \equiv (-1)^n \pmod{m}$. Умножая обе части данного сравнения на $(-1)^n b$, получаем:

$$aP_{n-1}(-1)^n b \equiv b \pmod{m}.$$

Таким образом, число $x = (-1)^n bP_{n-1}$ удовлетворяет сравнению (1.35). □

Пример 1.9. Требуется решить сравнение $9x \equiv 8 \pmod{34}$. Воспользуемся пунктом (*iii*) теоремы 1.31. Применяя алгоритм Евклида, получаем:

$$34 = 3 \cdot 9 + 7, \quad 9 = 1 \cdot 7 + 2, \quad 7 = 3 \cdot 2 + 1, \quad 2 = 2 \cdot 1.$$

Поэтому $\dfrac{34}{9} = 3 + 1\lrcorner 1 + 1\lrcorner 3 + 1\lrcorner 2$. Получаем:

n	-1	0	1	2	3
a_n	-	3	1	3	2
P_n	1	3	4	15	34

В данном случае $s = 3$, $P_{s-1} = 15$, $b = 8$. Тогда решение данного сравнения имеет вид:

$$x \equiv (-1)^3 \cdot 8 \cdot 15 = -120 \equiv -18 \equiv 16 \pmod{34}.$$

Теорема 1.32. Пусть $(a, m) = d$. Тогда если $d \nmid b$, то сравнение (1.35) не имеет решений. Если же $d|b$, то сравнение (1.35) имеет ровно d решений, которые принадлежат одному классу вычетов по модулю $\dfrac{m}{d}$.

Доказательство. Пусть $(a, m) = d$ и $d \nmid b$. Предположим, что в этом случае сравнение (1.35) имеет хотя бы одно решение x_0: $ax_0 \equiv b \pmod m$. Так как $d|a$ и $d|m$, то из пункта 9 теоремы 1.3 следует, что тогда $d|b$. Противоречие. Поэтому если $d \nmid b$, то сравнение (1.35) решений не имеет.

Покажем вторую часть теоремы. Пусть $d|b$. Тогда из пунктов 6 и 7 теоремы 1.26 будет следовать, что сравнение (1.35) будет эквивалентно сравнению $\widetilde{a}x \equiv \widetilde{b} \pmod{\widetilde{m}}$, где $\widetilde{a} = \dfrac{a}{d}$, $\widetilde{b} = \dfrac{b}{d}$, $\widetilde{m} = \dfrac{m}{d}$, причем $(\widetilde{a}, \widetilde{m}) = 1$. Из теоремы 1.30 следует, что сравнение $\widetilde{a}x \equiv \widetilde{b} \pmod{\widetilde{m}}$ имеет, и притом единственное, решение. Пусть x_0 — решение данного сравнения, причем $0 \leqslant x_0 < \widetilde{m}$ (напомним, что все элементы класса вычетов по модулю $\widetilde{m}$, порожденные элементом x_0, образуют одно решение). Рассмотрим элементы:

$$x_0,\ x_0 + \widetilde{m},\ x_0 + 2\widetilde{m}, \ldots,\ x_0 + (d-1)\widetilde{m}.$$

Данные элементы являются (одним) решением сравнения:

$$\widetilde{a}x \equiv \widetilde{b} \pmod{\widetilde{m}},$$

так как принадлежат одному классу по модулю $\widetilde{m}$, но все эти числа принадлежат попарно различным классам по модулю m, которые и образуют ровно d решений сравнения (1.35). □

1.13. Решение линейных диофантовых уравнений с использованием сравнений

В параграфе 1.5 приводился метод для отыскания общего решения систем диофантовых уравнений первой степени, в частности, для отыскания общего решения диофантовых уравнений

первой степени. В данном параграфе приведем метод нахождения частного решения диофантова уравнения первой степени с использованием сравнений.

Рассмотрим диофантово уравнение (см. параграф 1.4):

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b. \tag{1.36}$$

Опишем способ отыскания частного решения уравнения (1.36) на основе сравнений первой степени. При $n = 2$ частное решение находится с помощью обобщенного алгоритма Евклида. Поэтому пусть $n > 2$. Обозначим:

$$d_k = (a_k, a_{k+1}, \ldots, a_n), \quad k = 1, \ldots, n-1.$$

Будем полагать, что $d_1|b$. В этом случае уравнение (1.36) имеет решение (теорема 1.11). Зафиксируем некоторое $\widetilde{x}_1 \in \mathbb{Z}$ и запишем уравнение (1.36) в виде:

$$a_2x_2 + \ldots + a_nx_n = b - a_1\widetilde{x}_1. \tag{1.37}$$

Обозначим $b_1 = b - a_1\widetilde{x}_1$. Из теоремы 1.11 следует, что уравнение (1.37) имеет решение тогда и только тогда, когда $d_2|b_1$. Поэтому в качестве $\widetilde{x}_1$ возьмем любое целое число, удовлетворяющее сравнению:

$$a_1x \equiv b \pmod{d_2}.$$

Данное сравнение, учитывая теорему 1.30, имеет решение, так как $(a_1, d_2) = d_1|b$.

Если $n-1 = 2$, то для нахождения решения уравнения (1.37) применим обобщенный алгоритм Евклида. В противном случае, зафиксируем в качестве x_2 некоторое значение $\widetilde{x}_2$ и рассмотрим уравнение:

$$a_3x_3 + \ldots + a_nx_n = b_1 - a_2\widetilde{x}_2. \tag{1.38}$$

Обозначим $b_2 = b_1 - a_2\widetilde{x}_2$. Как и ранее, из теоремы 1.11 следует, что уравнение (1.38) имеет решение тогда и только тогда, когда $d_3|b_2$. Поэтому в качестве $\widetilde{x}_2$ возьмем любое целое число, удовлетворяющее сравнению:

$$a_2x \equiv b_1 \pmod{d_3}.$$

Данное сравнение имеет решение, так как $(a_2, d_3) = d_2 | b_1$.

И так далее. Предположим, что найдены значения $\widetilde{x}_1,\ldots,\widetilde{x}_{n-2}$. Для нахождения частного решения $\widetilde{x}_{n-1},\widetilde{x}_n$ уравнения:

$$a_{n-1}x_{n-1} + a_n x_n = b_{n-2} = b - a_1\widetilde{x}_1 - \ldots - a_{n-2}\widetilde{x}_{n-2}$$

применим обобщенный алгоритм Евклида. Тогда $\widetilde{x}_1,\ldots,\widetilde{x}_n$ — решение уравнения (1.36).

Пример 1.10. Найдем частное решение уравнения:

$$7x_1 + 6x_2 + 8x_3 + 12x_4 = 1.$$

Сначала найдем соответствующие наибольшие общие делители:

$$d_1 = (7, 6, 8, 12) = 1, \quad d_2 = (6, 8, 12) = 2, \quad d_3 = (8, 12) = 4.$$

Используя свойства наибольшего общего делителя, целесообразно сначала найти d_3, затем d_2 и, в последнюю очередь, d_1. В качестве $\widetilde{x}_1$ возьмем одно из решений сравнения:

$$7x \equiv 1 \pmod 2.$$

Пусть $\widetilde{x}_1 = 1$. Получим уравнение:

$$6x_2 + 8x_3 + 12x_4 = -6.$$

Рассмотрим сравнение:

$$6x \equiv -6 \equiv 2 \pmod 4.$$

Определим $\widetilde{x}_2 = 1$. Приходим к уравнению:

$$8x_3 + 12x_4 = -12,$$

для которого значения $\widetilde{x}_3 = 3$, $\widetilde{x}_4 = -3$ являются одним из решений.

Таким образом, $\widetilde{x}_1 = 1$, $\widetilde{x}_2 = 1$, $\widetilde{x}_3 = 3$, $\widetilde{x}_4 = -3$ — решение исходного уравнения.

1.14. Системы сравнений первой степени

Теорема 1.33 (**китайская теорема об остатках**). Пусть натуральные числа m_1, m_2,..., m_s являются попарно взаимно простыми. Тогда для любых целых чисел b_1, b_2,..., b_s система сравнений:

$$\begin{cases} x \equiv b_1 \pmod{m_1}, \\ \dots \\ x \equiv b_s \pmod{m_s} \end{cases} \tag{1.39}$$

имеет, и притом единственное, решение по модулю $m_1 \dots m_s$.

Доказательство. I способ. Покажем, что решением системы (1.39) является:

$$x_0 = \sum_{i=1}^{s} M_i N_i b_i,$$

где $M_i = m_1 \dots \widehat{m_i} \dots m_s$, $M_i N_i \equiv 1 \pmod{m_i}$, $i = 1, \dots, s$.

Так как для любого $i = 1, \dots, s$ числа m_i и M_i взаимно просты (предложение 1.6), то сравнение $M_i N_i \equiv 1 \pmod{m_i}$ имеет, и притом единственное, решение. При этом $M_i \equiv 0 \pmod{m_j}$, если $i \neq j$. Тогда для числа $x_0 = \sum_{i=1}^{s} M_i N_i b_i$ выполнены такие сравнения:

$$x_0 \equiv b_i \pmod{m_i}, \quad i = 1, \dots, s,$$

т.е. x_0 является решением системы (1.39).

Предположим, что y_0 — другое решение системы (1.39):

$$y_0 \equiv b_i \pmod{m_i}, \quad i = 1, \dots, s.$$

Тогда из теоремы 1.26 следует, что будут верны такие сравнения:

$$x_0 - y_0 \equiv 0 \pmod{m_i}, \quad i = 1, \dots, s.$$

Из той же теоремы следует, что:

$$x_0 - y_0 \equiv 0 \pmod{[m_1 \dots m_s]}.$$

Так как числа m_1,..., m_s попарно взаимно просты, то из теоремы 1.9 следует, что $[m_1 \dots m_s] = m_1 \dots m_s$. Поэтому:

$$x_0 - y_0 \equiv 0 \pmod{m_1 \dots m_s}.$$

Следовательно, x_0 и y_0 принадлежат одному и тому же классу вычетов по модулю числа $m_1 \dots m_s$.

II способ. Рассмотрим еще одно доказательство данной теоремы, которое пригодится при рассмотрении схемы разделения секрета на основе китайской теоремы об остатках.

Применим математическую индукцию по s. Пусть $s = 2$. Очевидно, что множество целых чисел, удовлетворяющих первому сравнению системы:

$$\begin{cases} x \equiv b_1 \pmod{m_1}, \\ x \equiv b_2 \pmod{m_2}, \end{cases} \tag{1.40}$$

имеет следующий вид: $x = b_1 + m_1 t$, $t \in \mathbb{Z}$. Поэтому второе сравнение будем решать в виде $b_1 + m_1 t \equiv b_2 \pmod{m_2}$ относительно неизвестной переменной t. Запишем последнее сравнение в виде $m_1 t \equiv b_2 - b_1 \pmod{m_2}$. Так как $(m_1, m_2) = 1$, то данное сравнение имеет, и притом единственное, решение t_0 (теорема 1.30). Таким образом, $x_0 = b_1 + m_1 t_0$ является решением системы (1.40). Единственность данного решения в диапазоне от 0 до $m_1 m_2 - 1$ показывается аналогично, как в первом способе доказательства.

Предположим, что утверждение теоремы верно для всех $k < s$, где $s \geqslant 3$. Исходя из данного предположения, пусть $\widetilde{x}$ — решение системы:

$$\begin{cases} x \equiv b_1 \pmod{m_1}, \\ \dots \\ x \equiv b_{s-1} \pmod{m_{s-1}}, \end{cases} \tag{1.41}$$

которое единственно в диапазоне от 0 до $m_1 \dots m_{s-1} - 1$, т.е. некоторое целое число x является решением системы (1.41) тогда и только тогда, когда $x \equiv \widetilde{x} \pmod{m_1 \dots m_{s-1}}$. Поэтому система (1.39) эквивалентна такой системе из двух сравнений:

$$\begin{cases} x \equiv \widetilde{x} \pmod{m_1 \dots m_{s-1}}, \\ x \equiv b_s \pmod{m_s}, \end{cases}$$

Так как числа $m_1 \dots m_{s-1}$ и m_s взаимно просты (предложение 1.6), то из базы индукции следует, что данная система имеет, и

притом единственное, решение в диапазоне от 0 до $m_1 \dots m_s - 1$. □

Следствие 1.8. Если числа $b_1, \dots, b_s$ в (1.39) независимо друг от друга пробегают полные системы вычетов по модулям $m_1, \dots, m_s$ соответственно, то числа x_0 — решения систем (1.39) пробегают полную систему вычетов по модулю $M = m_1 \dots m_s$.

Доказательство. Достаточно доказать, что получающиеся указанным в условии способом числа x_0 принадлежат различным классам вычетов по модулю M. Предположим, что для двух наборов $b_1, \dots, b_s$ и $\widetilde{b}_1, \dots, \widetilde{b}_s$ выполнено $b_j \not\equiv \widetilde{b}_j \pmod{m_j}$ для некоторого j. При этом предположим, что:

$$\sum_{i=1}^{s} M_i N_i b_i \equiv \sum_{i=1}^{s} M_i N_i \widetilde{b}_i \pmod{M}.$$

Так как $m_j \mid M$, то:

$$b_j \equiv M_j N_j b_j \equiv \sum_{i=1}^{s} M_i N_i b_i \equiv \sum_{i=1}^{s} M_i N_i \widetilde{b}_i \equiv M_j N_j \widetilde{b}_j \equiv \widetilde{b}_j \pmod{m_j}.$$

Противоречие. □

Следствие 1.9. Если числа $b_1, \dots, b_s$ в (1.39) независимо друг от друга пробегают приведенные системы вычетов по модулям $m_1, \dots, m_s$ соответственно, то числа x_0 — решения систем (1.39) пробегают приведенную систему вычетов по модулю $M = m_1 \dots m_s$.

Доказательство. Учитывая следствия 1.5 и 1.8, получаем:

$$(M, x_0) = (m_1 \dots m_s, x_0) \mid (m_1, x_0) \dots (m_s, x_0) =$$

$$= (m_1, b_1) \dots (m_s, b_s) = 1.$$

Поэтому $(M, x_0) = 1$. □

Следствие 1.10. Пусть $m_1, \dots, m_s$, $\widetilde{m}_1, \dots, \widetilde{m}_t$ — попарно взаимно простые натуральные числа и $b_1, \dots, b_s$, $\widetilde{b}_1, \dots, \widetilde{b}_t$ — произвольные целые числа. Тогда верны следующие утверждения.

(i) Если x_0 — решение системы (1.39), причем:

$$0 \leqslant x_0 < m_1 \dots m_s = M,$$

то решение системы:

$$\begin{cases} x \equiv b_1 \pmod{m_1}, \\ \dots \\ x \equiv b_s \pmod{m_s}, \\ x \equiv \widetilde{b}_1 \pmod{\widetilde{m}_1}, \\ \dots \\ x \equiv \widetilde{b}_t \pmod{\widetilde{m}_t} \end{cases}$$

в диапазоне от 0 до $m_1 \dots m_s \widetilde{m}_1 \dots \widetilde{m}_t - 1$ принадлежит множеству:

$$\{x_0 + i_1 M + i_2 M \widetilde{m}_1 + \dots + i_t M \widetilde{m}_1 \dots \widetilde{m}_{t-1} \mid \\ 0 \leqslant i_1 < \widetilde{m}_1, \dots, \ 0 \leqslant i_t < \widetilde{m}_t\}.$$

(ii) Рассмотрим систему (1.39). Пусть k — некоторое фиксированное целое число, причем $1 \leqslant k \leqslant s$. Рассмотрим все произведения вида $m_{i_1} \dots m_{i_k}$, каждое из которых является произведением k попарно различных элементов множества $\{m_1, \dots, m_s\}$. Обозначим через $\min(k)$ наименьшее из всех таких произведений:

$$\min(k) = \min_{1 \leqslant i_1 < \dots < i_k \leqslant s} m_{i_1} \dots m_{i_k}.$$

Пусть x_0 — решение системы (1.39), причем $0 \leqslant x_0 < \min(k)$. Тогда для нахождения x_0 достаточно решить систему, составленную из любых k сравнений системы (1.39).

1.15. Сравнения по простому модулю

Рассмотрим сравнения с одним неизвестным произвольной степени вида:

$$a_0 + a_1 x + \dots + a_n x^n \equiv 0 \pmod{p}, \tag{1.42}$$

где p — простое число.

Теорема 1.34. Если $p \nmid a_n$, то сравнение (1.42) эквивалентно сравнению с коэффициентом при старшем члене, равным единице.

Доказательство. Пусть $p \nmid a_n$. Тогда $(a_n, p) = 1$ и сравнение $a_n y \equiv 1 \pmod{p}$ имеет, и притом единственное, решение y_0 относительно неизвестного y (теорема 1.30). Так как $(y_0, p) = 1$, то из пунктов 4 и 5 теоремы 1.26 следует, что сравнения (1.42) и:

$$a_0 y_0 + (a_1 y_0)x + \ldots + (a_n y_0)x^n \equiv 0 \pmod{p}$$

эквивалентны. Из той же теоремы следует, что последнее сравнение и:

$$b_0 + b_1 x + \ldots + x^n \equiv 0 \pmod{p},$$

где $b_i \equiv a_i y_0 \pmod{p}$, $i = 0, 1, \ldots, n-1$, также эквивалентны. □

Например, следующие сравнения эквивалентны:

$$2x^3 + 3x + 1 \equiv 0 \pmod{5}, \quad x^3 + 4x + 3 \equiv 0 \pmod{5}.$$

Теорема 1.35. 1. Пусть $f(x)$ и $g(x)$ — некоторые многочлены с целыми коэффициентами. Тогда сравнения по простому модулю

$$f(x) \equiv 0 \pmod{p} \tag{1.43}$$

и:

$$f(x) - (x^p - x)g(x) \equiv 0 \pmod{p}$$

эквивалентны.

2. Если степень многочлена $f(x)$ больше или равна числу p, то сравнение (1.43) может быть заменено эквивалентным сравнением степени, меньшей числа p.

Доказательство. 1. Пусть для некоторого целого x_0 выполнено сравнение $f(x_0) \equiv 0 \pmod{p}$. Так как для любого целого x_0 выполнено $x_0^p - x_0 \equiv 0 \pmod{p}$ (теорема Ферма 1.28), то из теоремы 1.26 следует, что:

$$f(x_0) - (x_0^p - x_0)g(x_0) \equiv 0 \pmod{p}.$$

Обратно, пусть $f(x_0) - (x_0^p - x_0)g(x_0) \equiv 0 \pmod{p}$ для некоторого целого x_0. Из той же теоремы Ферма следует, что для любого

целого x_0 выполнено сравнение $(x_0^p - x_0)g(x_0) \equiv 0 \pmod{p}$. Поэтому $f(x_0) \equiv 0 \pmod{p}$ (теорема 1.26).

2. Пусть $\deg f(x) \geqslant p$. Так как число p простое, то существует, и притом единственное, разложение вида:

$$f(x) \equiv (x^p - x)g(x) + r(x) \pmod{p},$$

где $g(x)$ и $r(x)$ — многочлены с целыми коэффициентами, причем степень многочлена $r(x)$ строго меньше числа p. Из предыдущего пункта видно, что сравнения $f(x) \equiv 0 \pmod{p}$ и $r(x) \equiv 0 \pmod{p}$ эквивалентны. □

Лемма 1.9. Пусть $f(x)$, $g(x)$, $h(x)$ и $r(x)$ — многочлены с целыми коэффициентами, причем $f(x) = g(x)h(x) + r(x)$ и все коэффициенты многочлена $r(x)$ делятся на простое число p. Тогда любое решение сравнения (1.43) является решением по крайней мере одного из сравнений:

$$g(x) \equiv 0 \pmod{p}, \quad h(x) \equiv 0 \pmod{p}.$$

Доказательство. Пусть x_0 — некоторое решение сравнения (1.43). Тогда:

$$g(x_0)h(x_0) + r(x_0) \equiv 0 \pmod{p}.$$

Так как все коэффициенты $r(x)$ делятся на p, то для любого x_0 выполнено $r(x_0) \equiv 0 \pmod{p}$. Поэтому:

$$g(x_0)h(x_0) \equiv 0 \pmod{p}.$$

Из предложения 1.12 следует, что хотя бы одно из чисел $g(x_0)$ и $h(x_0)$ делится на p. □

Теорема 1.36. Пусть $f(x) = a_0 + a_1x + \ldots + a_nx^n$, $n = \deg f(x)$ и $p \nmid a_n$. Тогда сравнение (1.43) имеет не более чем n решений.

Доказательство. Применим метод математической индукции по n. Пусть $n = 1$. Так как $p \nmid a_1$, то $(a_1, p) = 1$ и сравнение $a_0 + a_1x \equiv 0 \pmod{p}$ имеет единственное решение (теорема 1.30). Предположим, что теорема верна для $n - 1$.

Пусть $f(x) = a_0 + a_1x + \ldots + a_nx^n$, $n = \deg f(x)$ и $p \nmid a_n$. Если множество решений сравнения $f(x) \equiv 0 \pmod{p}$ пусто,

то теорема доказана. Поэтому пусть x_0 — некоторое решение данного сравнения. В силу простоты числа p найдется такой многочлен $g(x)$ с целыми коэффициентами, что:

$$f(x) \equiv (x - x_0)g(x) + f(x_0) \pmod{p}.$$

Так как $p|f(x_0)$, то из леммы 1.9 следует, что все решения сравнения (1.43) находятся среди решений сравнений:

$$x - x_0 \equiv 0 \pmod{p}, \quad g(x) \equiv 0 \pmod{p}.$$

Сравнение $x - x_0 \equiv 0 \pmod{p}$ имеет единственное решение. Старший коэффициент многочлена $g(x)$ сравним с a_n по модулю p и $\deg\ g(x) = n - 1$, поэтому по предположению индукции сравнение $g(x) \equiv 0 \pmod{p}$ имеет не более $n - 1$ решений. □

Следствие 1.11. Для любого простого числа p выполнено сравнение:

$$x^p - x \equiv x(x-1)(x-2)\dots(x-p+1) \pmod{p}. \tag{1.44}$$

Доказательство. Рассмотрим многочлен:

$$f(x) = x^p - x - x(x-1)(x-2)\dots(x-p+1) \in \mathbb{Z}[x].$$

Предположим, что $f(x) \not\equiv 0 \pmod{p}$. Обозначим через n наибольшее целое число, такое, что коэффициент при x^n не делится на p. Так как степень многочлена $f(x)$ не превосходит числа $p - 1$, то $n \leqslant p - 1$.

С другой стороны, для любого целого a, $0 \leqslant a \leqslant p - 1$, выполнено (согласно малой теореме Ферма):

$$f(a) \equiv a^p - a \equiv 0 \pmod{p}.$$

Поэтому количество корней многочлена $f(x)$ равно p и должно выполняться неравенство $p \leqslant n$ (теорема 1.36). Противоречие с тем, что $n \leqslant p - 1$. □

Теорема 1.37. Пусть $f(x) = a_0 + a_1x + \ldots + a_nx^n$, $n = \deg\ f(x)$ и сравнение (1.43) имеет более чем n решений. Тогда $p|a_i$ для всех $i = 0, \dots, n$.

Доказательство. Пусть $n = 1$. Если $a_0 + a_1 x \equiv 0 \pmod{p}$ имеет более одного решения, то из теоремы 1.30 следует, что $p|a_1$, поэтому $(a_1, p) = p$ и $p|a_0$.

Предположим, что теорема верна для $n-1$. Пусть:

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n, \quad n = \deg\ f(x)$$

и сравнение $f(x) \equiv 0 \pmod{p}$ имеет более чем n решений. Тогда из теоремы 1.36 следует, что $p|a_n$ и для любого целого x выполнено $a_n x^n \equiv 0 \pmod{p}$. Тогда, с учетом теоремы 1.26:

$$a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} \equiv 0 \pmod{p}, \tag{1.45}$$

причем данное сравнение эквивалентно сравнению (1.43). Так как сравнение (1.45) имеет более чем $n-1$ решений, то по предположению индукции $p|a_i$, $i = 0, 1, \ldots, n-1$. □

Пусть:

$$f(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} + x^n, \quad \deg\ f(x) = n < p, \tag{1.46}$$

и:

$$x^{p-1} - 1 \equiv f(x) q(x) + r(x) \pmod{p}, \quad \deg\ r(x) < n. \tag{1.47}$$

Теорема 1.38. Пусть $p \nmid a_0$ для многочлена (1.46), где p — простое число. Сравнение $f(x) \equiv 0 \pmod{p}$ имеет n решений тогда и только тогда, когда все коэффициенты многочлена $r(x)$ из (1.47) делятся на p.

Доказательство. *Достаточность.* По теореме 1.36 сравнение $f(x) \equiv 0 \pmod{p}$ имеет не более n решений. Из теоремы Ферма следует, что сравнение $x^{p-1} - 1 \equiv 0 \pmod{p}$ имеет ровно $p-1$ решений, причем любое решение данного сравнения по лемме 1.9 является решением хотя бы одного из сравнений:

$$f(x) \equiv 0 \pmod{p}, \quad q(x) \equiv 0 \pmod{p}.$$

Пусть s и t — соответственно количества решений последних сравнений. Тогда $s + t \geqslant p - 1$. Так как $\deg\ q(x) = p - 1 - n$ и по теореме 1.36 сравнение $q(x) \equiv 0 \pmod{p}$ имеет не более $p - 1 - n$ решений, то:

$$s \geqslant p - 1 - t \geqslant p - 1 - (p - 1 - n) = n.$$

Поэтому $s = n$.

Необходимость. Пусть x_0 — некоторое фиксированное решение сравнения $f(x) \equiv 0 \pmod{p}$. Так как $p \nmid a_0$, то $p \nmid x_0$. Поэтому x_0 является решением сравнения $x^{p-1} - 1 \equiv 0 \pmod{p}$. Следовательно:

$$r(x_0) \equiv x_0^{p-1} - 1 - f(x_0)q(x_0) \equiv 0 \pmod{p},$$

т.е. x_0 является решением сравнения $r(x) \equiv 0 \pmod{p}$. Получается, что данное сравнение имеет не менее n решений, причем $\deg\ r(x) < n$. Исходя из теоремы 1.37, все коэффициенты многочлена $r(x)$ делятся на p. □

Теорема 1.39. Пусть $f(x), d(x)$, $x^p - x$ — многочлены с целочисленными коэффициентами, причем если их рассматривать над полем $GF(p)$, то $d(x)$ — наибольший общий делитель многочленов $f(x)$ и $x^p - x$. Тогда множества решений сравнений:

$$f(x) \equiv 0 \pmod{p} \qquad d(x) \equiv 0 \pmod{p}$$

одинаковы.

Доказательство. Так как $f(x) \equiv d(x)h(x) \pmod{p}$ для некоторого многочлена $h(x)$ с целыми коэффициентами, то любое решение сравнения $d(x) \equiv 0 \pmod{p}$ является решением сравнения $f(x) \equiv 0 \pmod{p}$.

Обратно. Подобно следствию 1.1 можно утверждать, что найдутся многочлены с целочисленными коэффициентами $u(x)$ и $v(x)$, для которых:

$$d(x) \equiv u(x)f(x) + v(x)(x^p - x) \pmod{p}.$$

Если для некоторого целого a выполнено $f(a) \equiv 0 \pmod{p}$, то, учитывая малую теорему Ферма, получаем:

$$d(a) \equiv u(a)f(a) + v(a)(a^p - a) \equiv 0 \pmod{p}.$$ □

Пример 1.11. Требуется решить сравнение:

$$x^3 - 3x^2 - 2x - 5 \equiv 0 \pmod{11}.$$

Учитывая теорему 1.39, вычислим наибольший общий делитель многочленов $x^3 - 3x^2 - 2x - 5$ и $x^{11} - x$ в поле $GF(11)$, т.е. по модулю 11. Имеем:

$x^3 \equiv 3x^2 + 2x + 5 \pmod{11}$,
$x^4 \equiv 3x^3 + 2x^2 + 5x \equiv 3(3x^2 + 2x + 5) + 2x^2 + 5x \equiv 4 \pmod{11}$,
$x^8 \equiv 4^2 \equiv 5 \pmod{11}$,
$x^{11} \equiv x^8 x^3 \equiv 4x^2 + 10x + 3 \pmod{11}$,
$x^{11} - x \equiv 4x^2 + 9x + 3 \pmod{11}$.

Используя алгоритм Евклида, получаем:

$$(x^3 - 3x^2 - 2x - 5, x^{11} - x) = (x^3 - 3x^2 - 2x - 5, 4x^2 + 9x + 3) =$$

$$= (4x^2 + 9x + 3, 7x + 1) = 7x + 1.$$

Значит сравнение $x^3 - 3x^2 - 2x - 5 \equiv 0 \pmod{11}$ имеет те же решения, что и сравнение $7x + 1 \equiv 0 \pmod{11}$. Последнее сравнение имеет решение $x \equiv 3 \pmod{11}$.

1.16. Сравнения по составному модулю

Пусть $m_1, \ldots, m_s$ — попарно взаимно простые натуральные числа.

Теорема 1.40. Пусть $f(x)$ — многочлен с целыми коэффициентами. Тогда сравнение:

$$f(x) \equiv 0 \pmod{m_1 \ldots m_s} \tag{1.48}$$

эквивалентно системе сравнений:

$$\begin{cases} f(x) \equiv 0 \pmod{m_1}, \\ \ldots \\ f(x) \equiv 0 \pmod{m_s}. \end{cases} \tag{1.49}$$

При этом если k_i — количество решений i-го сравнения системы (1.49), то сравнение (1.48) имеет $k_1 \ldots k_s$ различных решений по модулю $m_1 \ldots m_s$.

Доказательство. Первая часть теоремы следует из теорем 1.9 и 1.26. Покажем вторую часть теоремы. Пусть M_i — множество различных решений сравнения $f(x) \equiv 0 \pmod{m_i}$, где

$i = 1, \ldots, s$, M — множество различных решений сравнения $f(x) \equiv 0 \pmod{m_1 \ldots m_s}$. Построим отображение:

$$g : M_1 \times \ldots \times M_s \to M$$

следующим образом. Пусть $(x_1, \ldots, x_s) \in M_1 \times \ldots \times M_s$, т.е. $f(x_i) \equiv 0 \pmod{m_i}$, $i = 1, \ldots, s$. Из теоремы 1.33 следует, что система сравнений:

$$\begin{cases} x \equiv x_1 \pmod{m_1}, \\ \ldots \\ x \equiv x_s \pmod{m_s} \end{cases}$$

имеет единственное решение по модулю числа $m_1 \ldots m_s$. Обозначим это решение через x_0. Тогда положим $g(x_1, \ldots, x_s) = x_0$.

Пусть $y \in M$. Тогда $(y, \ldots, y) \in M_1 \times \ldots \times M_s$, поэтому g — сюрьекция. Пусть $(x_1, \ldots, x_s), (y_1, \ldots, y_s) \in M_1 \times \ldots \times M_s$, причем $x_j \not\equiv y_j \pmod{m_j}$ для некоторого j. Тогда:

$$g(x_1, \ldots, x_s) \not\equiv g(y_1, \ldots, y_s) \pmod{m_1 \ldots m_s},$$

поэтому g — инъекция. □

Следствие 1.12. Пусть $m = p_1^{\alpha_1} \ldots p_s^{\alpha_s}$ — каноническое разложение числа m. Тогда сравнение:

$$f(x) \equiv 0 \pmod{m}$$

эквивалентно системе сравнений:

$$\begin{cases} f(x) \equiv 0 \pmod{p_1^{\alpha_1}}, \\ \ldots \\ f(x) \equiv 0 \pmod{p_s^{\alpha_s}}. \end{cases}$$

Пример 1.12. Сравнение:

$$x^2 + 6x + 8 \equiv 0 \pmod{15}$$

эквивалентно системе сравнений:

$$\begin{cases} x^2 + 2 \equiv 0 \pmod{3}, \\ x^2 + x + 3 \equiv 0 \pmod{5}. \end{cases}$$

Первое сравнение системы имеет решения 1 и 2, второе сравнение системы — числа 1 и 3. Поэтому исходное сравнение имеет четыре решения. Эти решения находятся из систем сравнений:

$$\begin{cases} x \equiv 1 \pmod 3, \\ x \equiv 1 \pmod 5, \end{cases} \quad \begin{cases} x \equiv 1 \pmod 3, \\ x \equiv 3 \pmod 5, \end{cases}$$

$$\begin{cases} x \equiv 2 \pmod 3, \\ x \equiv 1 \pmod 5, \end{cases} \quad \begin{cases} x \equiv 2 \pmod 3, \\ x \equiv 3 \pmod 5. \end{cases}$$

Используя китайскую теорему об остатках, получаем такие решения исходного сравнения: 1, 13, 11, 8.

1.17. Степенные вычеты

Пусть $(a, m) = 1$. Рассмотрим последовательность:

$$a,\ a^2,\ a^3, \ldots \pmod m. \tag{1.50}$$

Из теоремы Эйлера (теорема 1.27) следует, что среди данных степеней существуют такие, что $a^k \equiv 1 \pmod m$.

Определение 1.26. *Показателем числа a по модулю m* называется наименьшее положительное k, при котором $a^k \equiv 1 \pmod m$. Показатель будем обозначать $P_m(a)$. Если модуль m фиксирован, то будем просто писать $P(a)$.

Из данного определения видно, что $a^{P(a)} \equiv 1 \pmod m$ и если $P(a) > 1$, то для любого $1 \leqslant r < P(a)$ выполнено условие $a^r \not\equiv 1 \pmod m$.

Теорема 1.41 (свойства показателя).

1. Если $a \equiv b \pmod m$, то $P(a) = P(b)$.
2. Если $a^n \equiv 1 \pmod m$, то $P(a) | n$.
3. $P(a) | \varphi(m)$, $P(a) | L(m)$, где φ — функция Эйлера, L — обобщенная функция Эйлера.
4. $a^s \equiv a^t \pmod m$ тогда и только тогда, когда $s \equiv t \pmod{P(a)}$.
5. Все элементы последовательности (1.50) принадлежат $P(a)$ различным классам эквивалентных элементов по модулю m, представителями которых являются числа a, a^2,..., $a^{P(a)}$.

6. $P(a^s) = P(a)$ тогда и только тогда, когда $(s, P(a)) = 1$.

7. Числа a, a^2,..., $a^{P(a)}$ являются различными решениями сравнения $x^{P(a)} \equiv 1 \pmod{m}$.

8. Если число $m = p$ простое, то числа a, a^2,..., $a^{P(a)}$ представляют собой все решения сравнения $x^{P(a)} \equiv 1 \pmod{p}$.

Доказательство. 1. Из теоремы 1.26 следует, что для любого натурального k выполнено сравнение $a^k \equiv b^k \pmod{m}$. Для определенности предположим, что $P(a) \leqslant P(b)$. Тогда:

$$b^{P(a)} \equiv a^{P(a)} \equiv 1 \pmod{m}.$$

Поэтому из определения показателя следует, что $P(b) \leqslant P(a)$. Таким образом, $P(a) = P(b)$.

2. Разложим число n по модулю числа $P(a)$: $n = qP(a) + r$, $0 \leqslant r < P(a)$. Тогда:

$$1 \equiv a^n = a^{qP(a)+r} = \left(a^{P(a)}\right)^q a^r \equiv a^r \pmod{m}.$$

Поэтому из определения числа $P(a)$ следует, что $r = 0$.

3. Следует из пункта 2 и теоремы Эйлера и 1.29.

4. Пусть $a^s \equiv a^t \pmod{m}$ и $s \leqslant t$. Так как $(a, m) = 1$, то в данном сравнении можно сокращать обе части на a (теорема 1.26). Получаем, что $1 \equiv a^{t-s} \pmod{m}$. Поэтому из пункта 2 следует, что $P(a)|(t - s)$, т.е. $s \equiv t \pmod{P(a)}$.

Обратно, пусть $s \equiv t \pmod{P(a)}$. Тогда для некоторого целого числа q выполнено равенство $s = t + qP(a)$ и:

$$a^s = a^{t+qP(a)} = a^t \left(a^{P(a)}\right)^q \equiv a^t \pmod{m}.$$

5. Для начала заметим, что числа a, a^2,..., $a^{P(a)}$ принадлежат $P(a)$ различным классам эквивалентных элементов по модулю m, так как если $a^s \equiv a^t \pmod{m}$ для некоторых s и t с условием $1 \leqslant s < t \leqslant P(a)$, то из пункта 4 следует, что $t - s$ делится на $P(a)$, что возможно только при $P(a) = 1$ и привело бы к противоречию с неравенством $1 \leqslant s < t \leqslant P(a)$.

Далее, рассмотрим элемент a^n последовательности (1.50). Так как $n = qP(a) + r$, $0 \leqslant r < P(a)$, то $a^n \equiv a^r \pmod{m}$.

6. Пусть $(s, P(a)) = 1$. Так как $1 \equiv (a^s)^{P(a^s)} = a^{sP(a^s)}$, то из пункта 2 следует, что $P(a)|sP(a^s)$. А из $(s, P(a)) = 1$ следует, что $P(a)|P(a^s)$ (предложение 1.5). При этом:

$$(a^s)^{P(a)} = \left(a^{P(a)}\right)^s \equiv 1 \pmod{m}.$$

Из пункта 2 следует, что $P(a^s)|P(a)$. Поэтому $P(a^s) = P(a)$.

Обратно, пусть $P(a^s) = P(a)$. Если $(s, P(a)) = d > 1$, то $s = xd$, $P(a) = yd$, где $0 < y < P(a)$. Тогда:

$$(a^s)^y = \left(a^{xd}\right)^y = \left(a^{yd}\right)^x = \left(a^{P(a)}\right)^x \equiv 1 \pmod{m}.$$

Поэтому $P(a^s) < P(a)$. Противоречие.

7. Следует из пункта 5.

8. Следует из пункта 7 и теоремы 1.36. □

Обозначим через $\psi(k)$ (при фиксированном модуле m) число классов, взаимно простых с m, для которых показатель равен k. Пусть $x_1, \ldots, x_{\varphi(m)}$ — приведенная система вычетов по модулю m. Тогда:

$$\psi(k) = \left|\{a \in \{x_1, \ldots, x_{\varphi(m)}\} \mid P(a) = k\}\right|.$$

Следующее свойство значений $\psi(k)$ следует из определения данных чисел.

Теорема 1.42.

$$\sum_{k \mid \varphi(m)} \psi(k) = \varphi(m).$$

Теорема 1.43. Для любого простого модуля p и любого натурального k верно равенство:

$$\psi(k) = \begin{cases} \varphi(k), & k \mid (p-1), \\ 0, & k \nmid (p-1). \end{cases}$$

Доказательство. Если $k \nmid (p-1) = \varphi(p)$, то понятно, что $\psi(k) = 0$. Поэтому пусть $k|(p-1)$.

Сначала покажем, что $\psi(k) \leqslant \varphi(k)$. Если $\psi(k) = 0$, то очевидно, что $\psi(k) < \varphi(k)$. Пусть $\psi(k) > 0$. Тогда для некоторого целого a будет выполнено равенство $P(a) = k$.

Рассмотрим сравнение $x^k \equiv 1 \pmod{p}$. Данное сравнение имеет не более k решений (теорема 1.36). При этом числа:

$$a,\ a^2, \ldots,\ a^k \tag{1.51}$$

являются k различными решениями данного сравнения (теорема 1.41). Это означает, что если некоторое целое число b имеет показатель k, то оно сравнимо с одним из чисел (1.51) по модулю p. А если некоторое целое число b не сравнимо ни с одним из чисел (1.51), то оно не может иметь показатель, равный k.

Некоторое число a^s из (1.51) имеет показатель k тогда и только тогда, когда $(s, k) = 1$ (теорема 1.41). Поэтому в данном случае $\psi(k) = \varphi(k)$.

Из следствия 1.7 и теоремы 1.42 следует, что:

$$\sum_{k \mid \varphi(p)} (\varphi(k) - \psi(k)) = 0.$$

При этом из неравенства $\psi(k) \leqslant \varphi(k)$ видно, что каждое слагаемое данной суммы неотрицательно. Поэтому данная сумма равна нулю только при условии, что $\psi(k) = \varphi(k)$. □

1.18. Первообразные корни по простому модулю

Определение 1.27. Целое число a называется *первообразным корнем по модулю* m, если $(a, m) = 1$ и выполнено равенство $P(a) = \varphi(m)$. В этом случае мультипликативная группа $\mathbb{Z}_m^*$ (кольца вычетов $\mathbb{Z}_m$) является циклической, а элемент a является образующим элементом группы $\mathbb{Z}_m^*$:

$$\mathbb{Z}_m^* = \{a, a^2, \ldots, a^{\varphi(m)} = 1\}.$$

Предложение 1.40. Если a — первообразный корень по модулю m, то числа:

$$a,\ a^2, \ldots,\ a^{\varphi(m)} \equiv 1 \pmod{m}$$

образуют приведенную систему вычетов по модулю m.

Доказательство следует из того, что данные элементы попарно несравнимы по модулю m (теорема 1.41), каждый из них взаимно прост с m и данных элементов ровно $\varphi(m)$.

Теорема 1.44. По простому модулю p существует $\varphi(p-1)$ первообразных корней.

Доказательство следует из теоремы 1.43.

Для нахождения первообразных корней можно использовать следующее утверждение.

Теорема 1.45. Пусть $a \in \mathbb{Z}$, $m \in \mathbb{N}$, $(a,m) = 1$, $\varphi(m) = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ — каноническое разложение числа $\varphi(m)$. Число a является первообразным корнем по модулю m тогда и только тогда, когда:

$$a^{\frac{\varphi(m)}{p_i}} \not\equiv 1 \pmod{m}, \quad i = 1, \dots, n.$$

Доказательство. Необходимое условие первообразного корня очевидно.

Покажем достаточное условие. Пусть некоторое целое число a не является первообразным корнем по модулю m. Тогда выполнено $P(a) < \varphi(m)$. Так как $P(a) | \varphi(m)$ (теорема 1.41), то, учитывая теорему 1.17:

$$P(a) = p_1^{\beta_1} p_2^{\beta_2} \dots p_n^{\beta_n}, \quad 0 \leqslant \beta_i \leqslant \alpha_i, \quad i = 1, \dots, n.$$

А из строгого неравенства $P(a) < \varphi(m)$ следует, что для некоторого i выполнено строгое неравенство $\beta_i < \alpha_i$. Это означает, что число:

$$p_1^{\alpha_1} \dots p_i^{\alpha_i - 1} \dots p_n^{\alpha_n} = \frac{\varphi(m)}{p_i}$$

делится на $P(a)$ и поэтому:

$$a^{\frac{\varphi(m)}{p_i}} \equiv 1 \pmod{m}. \qquad \square$$

Следствие 1.13. Пусть $p-1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n}$ — каноническое разложение числа $p-1$, где p — некоторое простое число. Число a является первообразным корнем по модулю p, $p \nmid a$, тогда и только тогда, когда:

$$a^{\frac{p-1}{p_i}} \not\equiv 1 \pmod{p}, \quad i = 1, \dots, n.$$

Пусть a — первообразный корень по модулю p. Всего имеется $\varphi(p-1)$ первообразных корней. Пусть:

$$M = \{s = 1,\ 2, \ldots, p-2 \mid (s, p-1) = 1\}.$$

Тогда $\{a^s \pmod p \mid s \in M\}$ — все первообразные корни по модулю p.

Алгоритм 1.3 (**вычисление первообразного корня**).
Вход: простое число p и разложение $p-1 = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$.
Выход: первообразный корень по модулю p.
1. Выбрать случайное число a, $1 < a < p$.
2. Цикл $i = 1, \ldots, n$
   Если $a^{\frac{p-1}{p_i}} \equiv 1 \pmod p$, то вернуться в шаг 1.
3. Вернуть значение a.

При случайном выборе вычета a вероятность того, что он окажется первообразным корнем по модулю p при известном разложении $p-1 = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_n^{\alpha_n}$, равна

$$\frac{\varphi(p-1)}{p-1} = \prod_{i=1}^{n} \left(1 - \frac{1}{p_i}\right) > \frac{1}{6 \ln \ln (p-1)},$$

где последнее неравенство следует из работы [33].

1.19. Первообразные корни по составному модулю

Рассмотрим вопрос о существовании первообразных корней по составному модулю.

Лемма 1.10. Пусть p — простое, $k, n \in \mathbb{N}$, $k \leqslant p^n$. Тогда $p^{n+1} \mid C_{p^n}^k p^k$. Если же $p > 2$, $k > 1$, то $p^{n+2} \mid C_{p^n}^k p^k$.

Доказательство. Так как:

$$C_{p^n}^k p^k = \frac{p^n \cdot (p^n - 1) \cdot \ldots \cdot (p^n - k + 1)}{k!} p^k,$$

то осталось заметить, что для показателя s наивысшей степени p, делящей $k!$, выполнено неравенство $s < k$, а при $p > 2$ и $k > 1$ выполнено $s + 1 < k$ (предложение 1.16). □

Предложение 1.41. Пусть p — простое, $n \in \mathbb{N}$. Если a — первообразный корень по модулю p^n, то a также является первообразным корнем по модулю p.

Доказательство. Пусть a — первообразный корень по модулю p^n. Тогда $P_{p^n}(a) = \varphi(p^n) = p^{n-1}(p-1)$.

Предположим, что a не является первообразным корнем по модулю p. Тогда для некоторого k, $1 \leqslant k < p-1$, выполнено $a^k \equiv 1 \pmod{p}$, что эквивалентно существованию такого целого t, для которого $a^k = 1 + pt$. Раскладывая $a^{kp^{n-1}} = (1+pt)^{p^{n-1}}$ по степеням pt, получаем:

$$a^{kp^{n-1}} = \sum_{i=0}^{p^{n-1}} C_{p^{n-1}}^i (pt)^i \equiv 1 \pmod{p^n},$$

так как все слагаемые с индексами $i = 1, 2, \ldots, p^{n-1}$ делятся на p^n (лемма 1.10). Так как $kp^{n-1} < \varphi(p^n)$, то a не является первообразным корнем по модулю p^n. Противоречие с условием утверждения. □

Таким образом, задача отыскания первообразных корней по модулю p^n сводится к тому, чтобы среди первообразных корней по модулю p отобрать числа, которые являются также первообразными корнями по модулю p^n.

Предложение 1.42. Если a — первообразный корень по простому модулю p и:

$$a^{p^{n-2}(p-1)} \not\equiv 1 \pmod{p^n}, \tag{1.52}$$

где $n \in \mathbb{N}$, $n \geqslant 2$, то a является также первообразным корнем по модулю p^n.

Доказательство. Пусть $P_{p^n}(a) = k$. Тогда, учитывая теорему 1.41, выполнено:

$$a^k \equiv 1 \pmod{p^n}, \qquad k \mid \varphi(p^n) = p^{n-1}(p-1).$$

Так как $p | p^n$, то $a^k \equiv 1 \pmod{p}$. Поэтому $p-1 \mid k$ (теорема 1.41) и для некоторого целого t выполнено $k = (p-1)t$. Учитывая $k \mid p^{n-1}(p-1)$, получаем:

$$(p-1)t \mid p^{n-1}(p-1), \quad t \mid p^{n-1}, \quad t = p^s, \quad 0 \leqslant s \leqslant n-1.$$

Следовательно, $k = p^s(p-1)$, $a^{p^s(p-1)} \equiv 1 \pmod{p^n}$.

Предположим, что $0 \leqslant s \leqslant n-2$. Возведем обе части сравнения:

$$a^{p^s(p-1)} \equiv 1 \pmod{p^n}$$

в степень p^{n-2-s} :

$$a^{p^{n-2}(p-1)} \equiv 1 \pmod{p^n}.$$

Получаем противоречие с (1.52). Поэтому:

$$s = n-1, \quad P_{p^n}(a) = k = p^{n-1}(p-1) = \varphi(p^n). \qquad \square$$

Следствие 1.14. Пусть $a \in \mathbb{Z}$, $n \in \mathbb{N}$, $n \geqslant 2$, p — простое, $p \nmid a$. a — первообразный корень по модулю p и выполнено (1.52) тогда и только тогда, когда a — первообразный корень по модулю p^n.

Предложение 1.43. Если a — первообразный корень по простому модулю $p > 2$, то из двух чисел a и $a+p$ по крайней мере одно является первообразным корнем по модулю p^2.

Доказательство. Так как $a+p$ принадлежит классу вычетов, порожденному элементом a, то $a+p$ также является первообразным корнем по модулю p, причем $p \nmid a$. Предположим, что a и $a+p$ не являются первообразными корнями по модулю p^2. Тогда по предложению 1.42:

$$a^{p-1} \equiv 1 \pmod{p^2}, \quad (a+p)^{p-1} \equiv 1 \pmod{p^2}.$$

Почленно вычитая правые и левые части данных сравнений, получаем:

$$(a+p)^{p-1} - a^{p-1} \equiv 0 \pmod{p^2}.$$

При этом все слагаемые правой части равенства:

$$(a+p)^{p-1} - a^{p-1} = \sum_{i=0}^{p-2} C_{p-1}^i a^i p^{p-1-i}$$

с номерами $i = 0, 1, \ldots, p-3$ делятся на p^2. Значит:

$$(a+p)^{p-1} - a^{p-1} \equiv (p-1)a^{p-2}p \pmod{p^2}.$$

Следовательно, $p^2 \mid (p-1)a^{p-2}p$, $p \mid (p-1)a^{p-2}$, что противоречит тому, что $p \nmid p-1$, $p \nmid a$. $\square$

Предложение 1.44. Если a — первообразный корень по модулю p^2, где p — нечетное простое число, то a является первообразным корнем по модулю p^n при любом $n \geqslant 1$.

Доказательство. Из предложения 1.41 следует, что a является первообразным корнем по модулю p. Так как:

$$P_{p^2}(a) = \varphi(p^2) = p(p-1),$$

то очевидно, что:

$$a^{p-1} \not\equiv 1 \pmod{p^2}. \tag{1.53}$$

Из $a^{p-1} \equiv 1 \pmod{p}$, $a^{p-1} = 1 + pt$ и (1.53) следует, что $p \nmid t$, поэтому $(p, t) = 1$. Методом математической индукции покажем, что для любого $n \in \mathbb{N}$ число a является первообразным корнем по модулю p^n. База индукции при $n = 1$, $n = 2$ очевидна.

Пусть $n \geqslant 3$, $P_{p^n}(a) = k$. Тогда:

$$a^k \equiv 1 \pmod{p^n} \equiv 1 \pmod{p^{n-1}}.$$

Из предположения индукции и теоремы 1.41 следует, что:

$$P_{p^{n-1}}(a) = \varphi(p^{n-1}) = p^{n-2}(p-1) \mid k.$$

С другой стороны, $k \mid \varphi(p^n) = p^{n-1}(p-1)$ (теорема 1.41). Поэтому либо $k = \varphi(p^{n-1})$, либо $k = \varphi(p^n)$.

Предположим, что $k = \varphi(p^{n-1})$: $a^{\varphi(p^{n-1})} \equiv 1 \pmod{p^n}$. Тогда, учитывая лемму 1.10, получаем:

$$1 \equiv a^{\varphi(p^{n-1})} = a^{p^{n-2}(p-1)} = (1+pt)^{p^{n-2}} \equiv 1 + tp^{n-1} \pmod{p^n},$$

что равносильно $tp^{n-1} \equiv 0 \pmod{p^n}$ или $p \mid t$. Противоречие с условием $(p, t) = 1$.

Таким образом, $P_{p^n}(a) = k = \varphi(p^n)$. □

Теорема 1.46. Если a — первообразный корень по простому нечетному модулю p и выполнено:

$$a^{p-1} \not\equiv 1 \pmod{p^2},$$

то a является первообразным корнем по модулю p^n при любом $n \in \mathbb{N}$.

Доказательство следует из предложений 1.42 и 1.44.

Теорема 1.47. Для любого модуля p^n, где p — нечетное простое число и $n \in \mathbb{N}$, существует не менее $\varphi(p-1)$ первообразных корней.

Доказательство. Все первообразные корни по модулю p^n содержатся среди первообразных корней по модулю p (предложение 1.41). По модулю p всего имеется $\varphi(p-1)$ первообразных корней (теорема 1.44). Пусть a — некоторый первообразный корень по модулю p. Тогда из предложений 1.43 и 1.44 следует, что либо a, либо $a+p$ является первообразным корнем по модулю p^n, причем $a \equiv a+p \pmod{p}$. □

Для нахождения первообразного корня по модулю p^n можно воспользоваться алгоритмом 1.3 и выбрать произвольный первообразный корень a по модулю p. Если для a выполнено $a^{p-1} \not\equiv 1 \pmod{p^2}$, то он и будет первообразным корнем по модулю p^n. В противном случае, первообразным корнем по модулю p^n будет $a+p$.

Теорема 1.48. По модулю $2p^n$, где p — нечетное простое число и $n \in \mathbb{N}$, существует не менее $\varphi(p-1)$ первообразных корней. Любой нечетный первообразный корень по модулю p^n является также первообразным корнем по модулю $2p^n$.

Доказательство. Для начала покажем, что для любого нечетного a и любого $n \in \mathbb{N}$:

$$a^k \equiv 1 \pmod{p^n} \Leftrightarrow a^k \equiv 1 \pmod{2p^n},$$

в частности $P_{p^n}(a) = P_{2p^n}(a)$. Действительно, из $a^k \equiv 1 \pmod{2p^n}$ следует $a^k \equiv 1 \pmod{p^n}$. Обратно, пусть $a^k \equiv 1 \pmod{p^n}$. Так как a — нечетное число, то $a \equiv 1 \pmod{2}$, поэтому выполнено $a^k \equiv 1 \pmod{2}$. Так как $(2, p^n) = 1$, $[2, p^n] = 2p^n$, то из теоремы 1.26 следует, что $a^k \equiv 1 \pmod{2p^n}$.

Пусть a — первообразный корень по модулю p^n (теорема 1.47). Тогда $a+p^n$ также является первообразным корнем по модулю p^n. При этом одно из чисел a и $a+p^n$ является нечетным. Обозначим его через b. Это число и есть первообразный корень по

модулю $2p^n$, так как:

$$P_{2p^n}(b) = P_{p^n}(b) = \varphi(p^n) = \varphi(2p^n). \qquad \square$$

Предложение 1.45. Для любого нечетного целого a и натурального $n \geqslant 3$ выполнено неравенство:

$$P_{2^n}(a) \leqslant 2^{n-2}.$$

Доказательство. Пусть $a = 1 + 2t$, $t \in \mathbb{Z}$. Применим математическую индукцию по n. При $n = 3$:

$$a^2 = (1 + 2t)^2 = 1 + 4t(t + 1) \equiv 1 \pmod{2^3},$$

т.е. $P_{2^3}(a) \leqslant 2$. Предположим, что утверждение выполнено для n: $P_{2^n}(a) \leqslant 2^{n-2}$. Так как $\varphi(2^n) = 2^{n-1}$ и $P_{2^n}(a) \mid 2^{n-1}$ (теорема 1.41), то $a^{2^{n-2}} \equiv 1 \pmod{2^n}$, что эквивалентно равенству:

$$a^{2^{n-2}} = 1 + 2^n s, \quad s \in \mathbb{Z}.$$

Возведем обе части данного равенства в квадрат:

$$a^{2^{n-1}} = (1 + 2^n s)^2 = 1 + 2^{n+1} s + 2^{2n} s^2 \equiv 1 \pmod{2^{n+1}},$$

т.е. $P_{2^{n+1}}(a) \leqslant 2^{n-1}$. $\square$

Поскольку $\varphi(2^n) = 2^{n-1}$, предложение показывает, что по модулю 2^n при $n \geqslant 3$ $P_{2^n}(a) < \varphi(2^n)$, т.е. по такому модулю не существует первообразных корней.

Теорема 1.49. Первообразные корни по модулю m существуют тогда и только тогда, когда $m \in \{1, 2, 4, p^n, 2p^n\}$, p — нечетное простое число, $n \in \mathbb{N}$.

Доказательство. Пусть $m \in \{1, 2, 4, p^n, 2p^n\}$. Для случаев $m = 1$, $m = 2$, $m = 4$ первообразными корнями будут соответственно числа 1, 1, 3. Для случаев $m = p^n$ и $m = 2p^n$ существование первообразных корней следует из теорем 1.47 и 1.48.

Обратно, пусть $m \notin \{1, 2, 4, p^n, 2p^n\}$. При $m = 2^n$, $n \geqslant 3$, первообразных корней нет (предложение 1.45).

Пусть m имеет хотя бы два различных нечетных простых делителя p и q. Тогда $p - 1$ и $q - 1$ не взаимно просты, поэтому

$L(m) < \varphi(m)$, где φ — функция Эйлера, L — обобщенная функция Эйлера. Также если $m = 2^k p^n$, $k \geqslant 2$, $n \geqslant 1$, то число 2^{k-1} и $p-1$ не взаимно просты, поэтому:

$$L(2^k p^n) = [2^{k-1}, p^{n-1}(p-1)] < \varphi(2^k p^n).$$

В обоих этих случаях $P_m(a) < \varphi(m)$ (теорема 1.41), поэтому по модулю m не существует первообразных корней. $\square$

1.20. Индексы (дискретные логарифмы)

Определение 1.28. Пусть $(a, m) = 1$, $(b, m) = 1$. Число s называется *индексом* (*дискретным логарифмом*) числа b по модулю m и основанию a, если:

$$a^s \equiv b \pmod{m}.$$

Число s обозначается $\operatorname{ind}_a b$ ($\log_a b$) или просто $\operatorname{ind} b$ при фиксированном основании a.

Пример 1.13. Пусть $m = 11$, $a = 2$. Так как $2^4 \equiv 5 \pmod{11}$, то $\operatorname{ind}_2 5 = 4$.

Из определения понятия индекса очевидным образом следует следующее предложение.

Предложение 1.46. Пусть $(a, m) = 1$, $(b, m) = 1$. Тогда:
1. $a^{\operatorname{ind}_a b} \equiv b \pmod{m}$.
2. если $a \equiv a_1 \pmod{m}$, $b \equiv b_1 \pmod{m}$, то $\operatorname{ind}_a b = \operatorname{ind}_{a_1} b_1$.

Пусть $s = \operatorname{ind}_a b$. Если $\widetilde{s}$ — другое число, для которого выполнено $a^{\widetilde{s}} \equiv b \pmod{m}$, то $s \equiv \widetilde{s} \pmod{P_m(a)}$ (теорема 1.41). Обратно, если $s \equiv \widetilde{s} \pmod{P_m(a)}$, причем s и $\widetilde{s}$ — неотрицательные целые числа, то $a^{\widetilde{s}} \equiv a^s \equiv b \pmod{m}$. Таким образом, множество индексов данного числа b образует подмножество в классе вычетов по модулю $P_m(a)$, состоящее из всех неотрицательных элементов данного класса.

В данном параграфе все первообразные корни будем обозначать буквой g.

Предложение 1.47. Пусть g — произвольный первообразный корень по модулю m. Тогда для любого b, $(b, m) = 1$, существуют индексы по основанию g, т.е. найдутся такие s, что $g^s \equiv b \pmod{m}$. Множество всех таких индексов для данного фиксированного числа b совпадает с неотрицательными числами некоторого класса вычетов по модулю $\varphi(m)$.

Доказательство. Согласно предложению 1.40 числа:

$$g,\ g^2, \ldots,\ g^{\varphi(m)} = 1 \pmod{m}$$

образуют приведенную систему вычетов по модулю m. Данную систему вычетов можно заменить на:

$$g^0 = 1,\ g, \ldots,\ g^{\varphi(m)-1} \pmod{m}.$$

Поэтому для любого целого числа b, $(b, m) = 1$, найдется такое $s \in \{0, 1, \ldots, \varphi(m) - 1\}$, что $g^s \equiv b \pmod{m}$, поэтому $s = \operatorname{ind}_g b$.

Вторая часть утверждения показана до данного предложения. □

Пример 1.14. Пусть $p = 11$. Число $g = 2$ является первообразным корнем по модулю 11. Составим таблицу степеней числа 2 по модулю 11:

s	0	1	2	3	4	5	6	7	8	9	10
$2^s \pmod{11}$	1	2	4	8	5	10	9	7	3	6	1

Тогда индексы чисел 1,2,. . . , 10 при основании $g = 2$ таковы:

a	1	2	3	4	5	6	7	8	9	10
$\operatorname{ind} a$	0	1	8	2	4	9	7	3	6	5

С помощью этой таблицы по данному числу a находится его индекс по основанию 2 и по модулю 11.

Следующая таблица позволяет по данному индексу находить соответствующее число:

$\operatorname{ind} a$	0	1	2	3	4	5	6	7	8	9	10
a	1	2	4	8	5	10	9	7	3	6	1

Определение 1.29. Если $\dfrac{a}{b} \equiv c \pmod m$, $(b, m) = 1$, то под $\operatorname{ind}_g \dfrac{a}{b}$ будем понимать $\operatorname{ind}_g c$.

Теорема 1.50 (**свойства индексов**)**.** Пусть g — первообразный корень по модулю m. Тогда верны следующие утверждения

1. Если $(a, m) = 1$, $(b, m) = 1$, то $a \equiv b \pmod m$ тогда и только тогда, когда $\operatorname{ind}_g a \equiv \operatorname{ind}_g b \pmod{\varphi(m)}$.

2. Если $(a, m) = 1$, $(b, m) = 1$, то:

$$\operatorname{ind}_g ab \equiv \operatorname{ind}_g a + \operatorname{ind}_g b \pmod{\varphi(m)}.$$

3. Если $(a_i, m) = 1$, $i = 1, \ldots, n$, то:

$$\operatorname{ind}_g a_1 \cdot \ldots \cdot a_n \equiv \operatorname{ind}_g a_1 + \ldots + \operatorname{ind}_g a_n \pmod{\varphi(m)}.$$

4. Если $(a, m) = 1$, то для любого натурального n выполнено:

$$\operatorname{ind}_g a^n \equiv n \cdot \operatorname{ind}_g a \pmod{\varphi(m)}.$$

5. Если $(a, m) = 1$, $(b, m) = 1$, то:

$$\operatorname{ind}_g \frac{a}{b} \equiv \operatorname{ind}_g a - \operatorname{ind}_g b \pmod{\varphi(m)}.$$

Доказательство. 1. Пусть $a \equiv b \pmod m$. Тогда согласно определению индекса:

$$g^{\operatorname{ind}_g a} \equiv a \equiv b \equiv g^{\operatorname{ind}_g b} \pmod m$$

и $\operatorname{ind}_g a \equiv \operatorname{ind}_g b \pmod{P_m(g) = \varphi(m)}$ (теорема 1.41).

Обратно, пусть $\operatorname{ind}_g a \equiv \operatorname{ind}_g b \pmod{\varphi(m)}$. Тогда для некоторого целого t выполнено $\operatorname{ind}_g a = \operatorname{ind}_g b + t\varphi(m)$ и, учитывая теорему Эйлера, получаем:

$$a \equiv g^{\operatorname{ind}_g a} = g^{\operatorname{ind}_g b} \left(g^{\varphi(m)}\right)^t \equiv g^{\operatorname{ind}_g b} \equiv b \pmod m.$$

2. По определению индексов чисел a и b имеем:

$$a \equiv g^{\operatorname{ind}_g a}, \quad b \equiv g^{\operatorname{ind}_g b},$$

$$g^{\operatorname{ind}_g ab} \equiv ab \equiv g^{\operatorname{ind}_g a} g^{\operatorname{ind}_g b} = g^{\operatorname{ind}_g a + \operatorname{ind}_g b} \pmod m.$$

Поэтому $\operatorname{ind}_g ab \equiv \operatorname{ind}_g a + \operatorname{ind}_g b \pmod{\varphi(m)}$.

3. Доказательство проходит методом математической индукции с учетом предыдущего пункта и того, что произведение любого непустого подмножества чисел из $\{a_1, \ldots, a_n\}$ взаимно просто с m.

4. Данное утверждение является частным случаем предыдущего утверждения.

5. Так как $(b, m) = 1$, то найдется такое целое число c, что $bc \equiv a \pmod{m}$ (теорема 1.30). Тогда, учитывая пункты 1 и 2, получаем:

$$\operatorname{ind}_g b + \operatorname{ind}_g c \equiv \operatorname{ind}_g bc \equiv \operatorname{ind}_g a \pmod{\varphi(m)}.$$

Так как $\frac{a}{b} = ab^{-1} \equiv cbb^{-1} \equiv c \pmod{m}$, то

$$\operatorname{ind}_g \frac{a}{b} \equiv \operatorname{ind}_g c \equiv \operatorname{ind}_g a - \operatorname{ind}_g b \pmod{\varphi(m)}. \qquad \square$$

Пример 1.15. 1. Пусть $p = 11$, $g = 2$. Тогда, учитывая пример 1.14:

$$\operatorname{ind} 9 = 6, \ \operatorname{ind} 7 = 7,$$
$$\operatorname{ind} 9 \cdot 7 = 6 + 7 \equiv 3 \pmod{10},$$
$$\operatorname{ind} \frac{9}{7} = 6 - 7 \equiv 9 \pmod{10}.$$

Поэтому:

$$9 \cdot 7 \equiv 8 \pmod{11}, \ \frac{9}{7} = 9 \cdot 7^{-1} \equiv 6 \pmod{11}.$$

Найдем $a \equiv 9^8 \pmod{11}$. Так как:

$$\operatorname{ind} a \equiv 8 \cdot \operatorname{ind} 9 \equiv 8 \cdot 6 \equiv 8 \pmod{10},$$

то $a \equiv 3 \pmod{11}$.

2. Найдем решение сравнения $8x \equiv 5 \pmod{11}$. Данное сравнение равносильно такому:

$$\operatorname{ind} 8 + \operatorname{ind} x \equiv \operatorname{ind} 5 \pmod{10}.$$

Так как $\operatorname{ind} x \equiv \operatorname{ind} 5 - \operatorname{ind} 8 = 1 \pmod{10}$, то $x \equiv 2 \pmod{11}$.

1.21. Сравнения второй степени

1.21.1. Квадратичные вычеты и невычеты

Рассмотрим сравнение вида:

$$x^n \equiv a(\mathrm{mod}\ m), \qquad (a, m) = 1. \tag{1.54}$$

Если сравнение (1.54) имеет решение, то число a называется *вычетом степени* n, в противном случае a называется *невычетом степени* n. При $n = 2$ вычеты или невычеты называются *квадратичными*. Рассмотрим сравнения второй степени по нечетному простому модулю:

$$x^2 \equiv a\ (\mathrm{mod}\ p), \quad (a, p) = 1, \quad p > 2. \tag{1.55}$$

Предложение 1.48. Если a — квадратичный вычет по модулю p, то сравнение (1.55) имеет ровно два решения.

Доказательство. Так как a — квадратичный вычет, то сравнение (1.55) имеет хотя бы одно решение: $x \equiv x_1\ (\mathrm{mod}\ p)$. Так как $(-x_1)^2 = x_1^2$, то сравнение (1.55) имеет и второе решение $x \equiv -x_1(\mathrm{mod}\ p)$, причем $x_1 \not\equiv -x_1(\mathrm{mod}\ p)$, так как, в противном случае из $x_1 \equiv -x_1\ (\mathrm{mod}\ p)$ выполнялось бы сравнение $2x_1 \equiv 0(\mathrm{mod}\ p)$. Из предложения 1.5 следует, что из $p|2x_1$ и $(p, x_1) = 1$ выполнено $p|2$, что неверно, так как $p > 2$.

Осталось заметить, что по теореме 1.36 сравнение (1.55) не может иметь более двух решений. □

Теорема 1.51 (критерий Эйлера). a — квадратичный вычет по простому модулю p, $p > 2$, $p \nmid a$, тогда и только тогда, когда выполнено сравнение:

$$a^{\frac{p-1}{2}} \equiv 1\ (\mathrm{mod}\ p). \tag{1.56}$$

Доказательство. I способ. Из теоремы 1.38 следует, что необходимым и достаточным условием того, чтобы сравнение (1.55) имело два решения, является делимость на p всех коэффициентов остатка от деления $x^{p-1} - 1$ на $x^2 - a$. Найдем этот остаток.

По теореме Безу $f(y) = (y-a)q(y) + f(a)$. Поэтому:

$$y^{\frac{p-1}{2}} - 1 = (y-a)q(y) + a^{\frac{p-1}{2}} - 1, \quad q(y) \in \mathbb{Z}[y].$$

При подстановке $y \to x^2$ в последнем равенстве получим:

$$x^{p-1} - 1 = (x^2 - a)q(x^2) + a^{\frac{p-1}{2}} - 1.$$

Поэтому необходимым и достаточным условием того, чтобы сравнение (1.55) имело два решения, является делимость $a^{\frac{p-1}{2}} - 1$ на p.

II способ. Пусть g — некоторый первообразный корень по модулю p (теорема 1.49). Для некоторых $y, k \in \mathbb{Z}$ выполнено $x = g^y$, $a = g^k \pmod p$. Подставим эти равенства в (1.55):

$$g^{2y} \equiv g^k \pmod p.$$

Так как $(p, g) = 1$, то данное сравнение эквивалентно сравнению $g^{2y-k} \equiv 1 \pmod p$ (теорема 1.26). Так как $P(g) = p - 1$, то из теоремы 1.41 следует, что последнее сравнение выполнено тогда и только тогда, когда $p-1 \mid 2y-k$, что равносильно сравнению $2y \equiv k \pmod{p-1}$. При этом последнее сравнение имеет решение относительно y тогда и только тогда, когда k — четное число, при этом решений будет ровно два (теорема 1.32). Значит из чисел приведенной системы вычетов по модулю p :

$$g^0,\ g^1, \ldots,\ g^{p-2}$$

квадратичными вычетами будут $g^0,\ g^2, \ldots,\ g^{p-3}$, а числа g^1, $g^3, \ldots, g^{p-2}$ — квадратичными невычетами. Итак, a — квадратичный вычет по модулю p тогда и только тогда, когда $a = g^{2t}$; a — квадратичный невычет по модулю p тогда и только тогда, когда $a = g^{1+2t}$.

Пусть $a = g^{2t}$. Тогда, учитывая малую теорему Ферма:

$$a^{\frac{p-1}{2}} = \left(g^{p-1}\right)^t \equiv 1 \pmod p.$$

Обратно, пусть выполнено сравнение (1.56), но a не является квадратичным вычетом. Тогда $a = g^{1+2t}$ и, учитывая малую теорему Ферма, получаем:

$$1 \equiv a^{\frac{p-1}{2}} = \left(g^{1+2t}\right)^{\frac{p-1}{2}} = g^{\frac{p-1}{2}} \left(g^{p-1}\right)^t \equiv g^{\frac{p-1}{2}} \pmod p.$$

Противоречие с тем, что g — первообразный корень по модулю p. $\square$

Теорема 1.52. a — квадратичный невычет по простому модулю p, $p > 2$, $p \nmid a$, тогда и только тогда, когда выполнено сравнение:

$$a^{\frac{p-1}{2}} \equiv -1 \pmod{p}. \tag{1.57}$$

Доказательство. Если $p \nmid a$, то по малой теореме Ферма имеем:

$$\left(a^{\frac{p-1}{2}} - 1\right)\left(a^{\frac{p-1}{2}} + 1\right) = a^{p-1} - 1 \equiv 0 \pmod{p}. \tag{1.58}$$

Если a — квадратичный невычет по модулю p, то первый множитель в (1.58) не может делиться на p (теорема 1.51). Тогда по предложению 1.12 на p делится второй множитель, т.е. выполнено условие (1.57).

Обратно. пусть выполнено условие (1.57). Заметим, что оба множителя в (1.58) не могут одновременно делиться на p, так как иначе их разность 2 тоже делилась бы на $p > 2$, что невозможно. Поэтому если выполнено условие (1.57), то в этом случае условие (1.56) не выполнено, поэтому a — квадратичный невычет. $\square$

Пример 1.16. 1. Число 2 — квадратичный вычет по модулю 7, так как $2^3 = 8 \equiv 1 \pmod{7}$.

2. Число 3 — квадратичный невычет по модулю 7, так как $3^3 = 27 \equiv -1 \pmod{7}$.

Теорема 1.53. По любому простому модулю $p > 2$ число классов квадратичных вычетов равно числу классов квадратичных невычетов, причем числа:

$$1^2, \quad 2^2, \quad \ldots, \left(\frac{p-1}{2}\right)^2, \tag{1.59}$$

образуют систему представителей всех классов квадратичных вычетов по модулю p.

Доказательство. Сначала покажем, что число всех классов квадратичных вычетов по модулю p равно $\dfrac{p-1}{2}$. Из теоремы

1.51 следует, что классы квадратичных вычетов по модулю p представляют собой решения сравнения:

$$x^{\frac{p-1}{2}} \equiv 1 (\mathrm{mod}\ p). \tag{1.60}$$

Поэтому число квадратичных вычетов равно числу решений этого сравнения. Так как:

$$x^{p-1} - 1 = \left(x^{\frac{p-1}{2}} - 1\right)\left(x^{\frac{p-1}{2}} + 1\right),$$

то из теоремы 1.38 следует, что сравнение (1.60) имеет ровно $\dfrac{p-1}{2}$ решений.

Всего по модулю p существует $p-1$ классов вычетов, взаимно простых с модулем, поэтому число квадратичных невычетов равно:

$$p - 1 - \frac{p-1}{2} = \frac{p-1}{2}.$$

Покажем вторую часть теоремы. Рассмотрим приведенную систему вычетов по модулю p :

$$-\frac{p-1}{2}, \ldots, -2,\ -1,\ 1,\ 2, \ldots, \frac{p-1}{2}.$$

Каждое из чисел (1.59) представляет собой квадратичный вычет по модулю p, так как сравнение $x^2 \equiv s^2 \pmod{p}$, где $1 \leqslant s \leqslant \dfrac{p-1}{2}$, имеет ровно два таких решения: $x \equiv \pm s \pmod{p}$. Все эти числа попарно несравнимы по модулю p, так как из $s^2 \equiv t^2 \pmod{p}$, $1 \leqslant s < t \leqslant \dfrac{p-1}{2}$, следовало бы, что сравнение $x^2 \equiv s^2 \pmod{p}$ имело бы четыре решения $(x = -t, -s, s, t)$ вопреки предложению 1.48. □

Пример 1.17. Квадратичные вычеты по модулю $p = 11$:

$$1^1 = 1,\ 2^2 = 4,\ 3^2 = 9,\ 4^2 = 16 \equiv 5 \pmod{11},\ 5^5 \equiv 3 \pmod{11}.$$

Классами квадратичных вычетов по модулю 11 являются классы: $\overline{1}$, $\overline{3}$, $\overline{4}$, $\overline{5}$, $\overline{9}$.

1.21.2. Символ Лежандра

Введем в рассмотрение такое понятие, как *символ Лежандра* $\left(\frac{a}{p}\right)$ (читается символ a по отношению к p, p — простое, $p > 2$, $p \nmid a$). Этот символ равен 1, если a — квадратичный вычет по модулю p, и -1, если a — квадратичный невычет.

Другими словами, $\left(\frac{a}{p}\right)$ равно 1, если сравнение $x^2 \equiv a \pmod{p}$ имеет (с учетом предложения 1.48) два решения, и -1, если это сравнение не имеет решений.

Например, $\left(\frac{3}{11}\right) = 1$, $\left(\frac{7}{11}\right) = -1$, что следует из примера 1.17.

Теорема 1.54 (**свойства символа Лежандра**).

1. Если $a \equiv b \pmod{p}$, то $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.
2. $\left(\frac{a^2}{p}\right) = 1$, в частности, $\left(\frac{1}{p}\right) = 1$.
3. $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$.
4. $\left(\frac{-1}{p}\right) = \begin{cases} 1, & p \equiv 1 \pmod 4, \\ -1, & p \equiv 3 \pmod 4. \end{cases}$
5. $\left(\frac{a_1 \dots a_n}{p}\right) = \left(\frac{a_1}{p}\right) \dots \left(\frac{a_n}{p}\right)$, $p \nmid a_i$, $i = 1, \dots, n$.

Доказательство. 1. Очевидно.

2. Сравнение $x^2 \equiv a^2 \pmod{p}$ имеет два решения, а именно, $x \equiv \pm a \pmod{p}$.

3. Следует из теорем 1.51 и 1.52.

4. Из пункта 3 следует, что $\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$. Если для некоторого целого m выполнено равенство $p = 4m + 1$, то

число $\dfrac{p-1}{2}$ является четным. Если же $p = 4m+3$, то число $\dfrac{p-1}{2}$ нечетно.

5. Учитывая пункт 3, имеем:

$$\left(\frac{a_1 \dots a_n}{p}\right) \equiv (a_1 \dots a_n)^{\frac{p-1}{2}} = a_1^{\frac{p-1}{2}} \dots a_n^{\frac{p-1}{2}} \equiv$$

$$\equiv \left(\frac{a_1}{p}\right) \dots \left(\frac{a_n}{p}\right) \pmod{p}.$$

Так как $p > 2$, а символ Лежандра принимает значения -1 или 1, то $\left(\dfrac{a_1 \dots a_n}{p}\right) = \left(\dfrac{a_1}{p}\right) \dots \left(\dfrac{a_n}{p}\right)$. □

Из теоремы 1.54 следует, что вычисление любого символа Лежандра сводится к вычислению лишь символов вида $\left(\dfrac{-1}{p}\right)$, $\left(\dfrac{2}{p}\right)$, $\left(\dfrac{q}{p}\right)$, где q — нечетное простое число, $q < p$.

Следующий критерий, отличный от критерия Эйлера, дает другой способ выяснить, является ли некоторое число a квадратичным вычетом или невычетом по модулю p.

Теорема 1.55 (**критерий Гаусса**)**.** Пусть p — простое нечетное число, a — целое число, не делящееся на p. Тогда:

$$\left(\frac{a}{p}\right) = (-1)^n,$$

где n — число отрицательных чисел из приведенной системы вычетов:

$$-\frac{p-1}{2}, \dots, -2,\ -1,\ 1,\ 2, \dots, \frac{p-1}{2}, \tag{1.61}$$

сравнимых с числами:

$$a,\ 2a, \dots,\ \frac{p-1}{2}a. \tag{1.62}$$

Доказательство. Заметим, что числа (1.62) попарно несравнимы по модулю p (лемма 1.8) и взаимно просты с p. Поэтому каждое из чисел (1.62) сравнимо с одним и только одним из чисел (1.61), так что каждому числу ia из (1.62) сопоставим число из (1.61) вида $(-1)^{t_i} r_i$, такое, что:

$$ia \equiv (-1)^{t_i} r_i \pmod{p}, \quad r_i \in \left\{1, 2, \ldots, \frac{p-1}{2}\right\}, \quad t_i \in \{0, 1\}. \tag{1.63}$$

При этом $r_i \neq r_j$ при $i \neq j$. Действительно, если $r_i = r_j$, то:

$$(-1)^{t_i} ia \equiv r_i \equiv r_j \equiv (-1)^{t_j} ja \pmod{p}.$$

Поэтому либо $ia \equiv ja \pmod{p}$, либо $ia \equiv -ja \pmod{p}$. Первое сравнение невозможно, так как из него следовало бы $i = j$. Второе сравнение означает:

$$i \equiv -j \pmod{p}, \quad p \,|\, (i + j), \quad 1 \leqslant i < j \leqslant \frac{p-1}{2},$$

что невозможно.

Перемножив сравнения (1.63) для всех $i = 1, 2, \ldots, \dfrac{p-1}{2}$, а затем сократив на число $1 \cdot 2 \cdot \ldots \cdot \frac{p-1}{2}$, взаимно простое с p, получим:

$$a^{\frac{p-1}{2}} \equiv (-1)^{t_1 + \ldots + t_{\frac{p-1}{2}}} = (-1)^n \pmod{p}.$$

Учитывая теорему 1.54, получаем $\left(\dfrac{a}{p}\right) \equiv (-1)^n \pmod{p}$. Так как $p > 2$, то последнее сравнение означает, что $\left(\dfrac{a}{p}\right) = (-1)^n$.

□

Следствие 1.15. 1. $\left(\dfrac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$.

2. Если a нечетно, то $\left(\dfrac{a}{p}\right) = (-1)^S$, где $S = \sum_{x=1}^{\frac{p-1}{2}} \left[\dfrac{ax}{p}\right]$.

Доказательство. Сохраним все обозначения предыдущей теоремы. Из сравнения $ak \equiv (-1)^{t_k} r_k \pmod{p}$, следует, что для

некоторого целого t выполнено равенство $ak = (-1)^{t_k} r_k + pt$, что равносильно $ak - pt = (-1)^{t_k} r_k$. Неполное частное от деления ak на p равно $\left[\dfrac{ak}{p}\right]$. Поэтому:

$$ak - p\left[\frac{ak}{p}\right] = \begin{cases} r_k, & t_k = 0, \\ p - r_k, & t_k = 1. \end{cases}$$

Просуммируем эти равенства для всех $k = 1, 2, \ldots, \dfrac{p-1}{2}$. Учитывая, что:

$$1 + 2 + \ldots + \frac{p-1}{2} = \frac{p^2 - 1}{8},$$

получим:

$$\frac{p^2-1}{8} \cdot a - p \cdot S = n \cdot p + \sum_{k=1}^{\frac{p-1}{2}} (-1)^{t_k} r_k.$$

Перейдем в этом равенстве к сравнению по модулю 2. При этом заметим, что:

$$\sum_{k=1}^{\frac{p-1}{2}} (-1)^{t_k} r_k \equiv \sum_{k=1}^{\frac{p-1}{2}} r_k = \sum_{k=1}^{\frac{p-1}{2}} k = \frac{p^2-1}{8} \pmod 2, \quad p \equiv \pm 1 \pmod 2.$$

Получаем:

$$\frac{p^2-1}{8} \cdot a + S \equiv \frac{p^2-1}{8} \cdot a - p \cdot S \equiv n + \frac{p^2-1}{8} \pmod 2,$$

$$n \equiv S + (a-1)\frac{p^2-1}{8} \pmod 2. \tag{1.64}$$

При $a = 2$ все слагаемые, представляющие S, равны нулю. Поэтому $S = 0$ и $n \equiv \dfrac{p^2-1}{8} \pmod 2$, что доказывает первый пункт следствия.

Если a — нечетно, то из (1.64) следует, что $n \equiv S \pmod 2$, что доказывает второй пункт следствия. $\square$

Замечание 1.5. $\dfrac{p^2-1}{8}$ четно, если $p=8k+1$ или $p=8k+7$, и нечетно, если $=8k+3$ или $p=8k+5$. Поэтому:

$$\left(\frac{2}{p}\right)=\begin{cases}1, & p\equiv 1 \pmod 8 \text{ или } p\equiv 7 \pmod 8,\\ -1, & p\equiv 3 \pmod 8 \text{ или } p\equiv 5 \pmod 8.\end{cases}$$

Пример 1.18. Определим, имеет ли квадратичное сравнение $x^2-2\equiv 0 \pmod{97}$ решение. Так как $97=8\cdot 12+1$, то $\left(\dfrac{2}{97}\right)=1$, поэтому сравнение имеет два решения.

Теорема 1.56 (квадратичный закон взаимности Гаусса). Для любых различных нечетных простых чисел p и q выполняется равенство

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Иначе:

$$\left(\frac{p}{q}\right)=\begin{cases}\left(\dfrac{q}{p}\right), & p\equiv 1 \pmod 4 \quad q\equiv 1 \pmod 4,\\ -\left(\dfrac{q}{p}\right), & p\equiv 3 \pmod 4 \quad q\equiv 3 \pmod 4.\end{cases}$$

Доказательство. По следствию 1.15 выполняются равенства $\left(\dfrac{q}{p}\right)=(-1)^{S_1}$, $\left(\dfrac{p}{q}\right)=(-1)^{S_2}$, где:

$$S_1=\sum_{x=1}^{\frac{p-1}{2}}\left[\frac{qx}{p}\right], \quad S_2=\sum_{y=1}^{\frac{q-1}{2}}\left[\frac{py}{q}\right].$$

Поэтому:

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{S_1+S_2}$$

и достаточно доказать равенство $S_1+S_2=\dfrac{p-1}{2}\cdot\dfrac{q-1}{2}$.

При фиксированном x из интервала $0 < x \leqslant \dfrac{p-1}{2}$ величина $\left[\dfrac{qx}{p}\right]$ есть количество целых точек y на промежутке $0 < y \leqslant \dfrac{qx}{p}$. Так как $\dfrac{qx}{p}$ не является целым числом, то величина $\left[\dfrac{qx}{p}\right]$ есть количество целых точек y на промежутке $0 < y < \dfrac{qx}{p}$. Поэтому величина S_1 равна количеству пар точек (x, y) с целочисленными координатами на координатной плоскости, удовлетворяющих неравенствам:

$$0 < x \leqslant \frac{p-1}{2}, \quad 0 < y < \frac{qx}{p}.$$

Так как $\dfrac{p}{2}$ не целое число и $\dfrac{qx}{p} < \dfrac{qp}{p2} = \dfrac{q}{2}$, то S_1 — количество точек $(x, y) \in \mathbb{Z}^2$ с условиями:

$$0 < x < \frac{p}{2}, \quad 0 < y < \frac{q}{2}, \quad py - qx < 0.$$

Аналогично, S_2 — количество точек $(x, y) \in \mathbb{Z}^2$ с условиями:

$$0 < x < \frac{p}{2}, \quad 0 < y < \frac{q}{2}, \quad py - qx > 0.$$

Пусть $T = \left\{(x, y) \in \mathbb{Z}^2 \mid 0 < x < \dfrac{p}{2},\ 0 < y < \dfrac{q}{2}\right\}$. Тогда $S_1 + S_2$ — количество элементов множества T с условием $py - qx \neq 0$. Но для любого элемента $(x, y) \in T$ выполнено $py - qx \neq 0$. Поэтому:

$$S_1 + S_2 = |T| = \frac{p-1}{2} \cdot \frac{q-1}{2}. \qquad \square$$

Приведенные свойства символа Лежандра позволяют указать алгоритм вычисления символа $\left(\dfrac{a}{p}\right)$.

Алгоритм 1.4 (вычисление символа Лежандра).
Вход: $a \in \mathbb{Z}$, p — нечетное простое.
Выход: $\left(\frac{a}{p}\right)$.

1. Заменить a на такое b, что $a \equiv b \pmod{p}$ и $|b| < \frac{p}{2}$.
2. Найти разложение числа b :

$$b = (-1)^s 2^t p_1^{k_1} \dots p_n^{k_n}.$$

Тогда:

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = \left(\frac{-1}{p}\right)^s \left(\frac{2}{p}\right)^t \left(\frac{p_1}{p}\right)^{k_1} \dots \left(\frac{p_n}{p}\right)^{k_n}.$$

3. Вычислить $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$, используя теорему 1.54 и следствие 1.15.

4. Символы $\left(\frac{p_i}{p}\right)$, $i = 1, \dots, n$, выразить через $\left(\frac{p}{p_i}\right)$, используя квадратичный закон взаимности. Далее для вычисления $\left(\frac{p}{p_i}\right)$ снова перейти к шагу 1.

Пример 1.19. Выясним, имеет ли решение квадратичное сравнение $x^2 - 42 \equiv 0 \pmod{107}$:

$$\left(\frac{42}{107}\right) = \left(\frac{2}{107}\right)\left(\frac{3}{107}\right)\left(\frac{7}{107}\right) =$$

$$= -\left(\frac{107}{3}\right)\left(\frac{107}{7}\right) = -\left(\frac{-1}{3}\right)\left(\frac{2}{7}\right) = 1.$$

Поэтому сравнение имеет два решения.

Основной сложностью в алгоритме 1.4 является факторизация числа b. Чтобы избавиться от необходимости факторизации целых чисел, вводится в рассмотрение символ Якоби.

1.21.3. Символ Якоби

Обобщением символа Лежандра является символ, введенный Якоби. Пусть m — нечетное, большее единицы, и $m = p_1^{k_1} \dots p_n^{k_n}$ — каноническое разложение числа m, $(a, m) = 1$. *Символ Якоби* $\left(\frac{a}{m}\right)$ определяется равенством:

$$\left(\frac{a}{m}\right) = \left(\frac{a}{p_1}\right)^{k_1} \dots \left(\frac{a}{p_n}\right)^{k_n},$$

где $\left(\frac{a}{p_i}\right)$ — символ Лежандра, $i = 1, \dots, n$.

Предложение 1.49. Для любых нечетных чисел $a_1, \dots, a_n$ выполнено сравнение:

$$(a_1 - 1) + \dots + (a_n - 1) \equiv a_1 \dots a_n - 1 \pmod 4,$$

что эквивалентно существованию такого целого t, для которого выполнено равенство:

$$\frac{a_1 - 1}{2} + \dots + \frac{a_n - 1}{2} = \frac{a_1 \dots a_n - 1}{2} + 2t.$$

Доказательство. Применим индукцию по n. При $n = 1$ утверждение очевидно. При $n = 2$ требуется показать, что:

$$a_1 + a_2 - 2 \equiv a_1 a_2 - 1 \pmod 4.$$

Так как $a_1 \equiv 1 \pmod 4$ или $a_1 \equiv 3 \pmod 4$ и $a_2 \equiv 1 \pmod 4$ или $a_2 \equiv 3 \pmod 4$, то проверяемое сравнение доказывается перебором всех данных вариантов.

При $n > 2$ положим $b = a_1 \dots a_{n-1}$, причем b — нечетное. По предположению индукции:

$$(a_1 - 1) + \dots + (a_{n-1} - 1) \equiv b - 1 \pmod 4.$$

Тогда:

$$(a_1 - 1) + \dots + (a_n - 1) = \sum_{i=1}^{n-1} (a_i - 1) + (a_n - 1) \equiv$$

$$\equiv (b-1)+(a_n-1) \equiv ba_n - 1 \equiv a_1 \dots a_n - 1 \pmod 4. \qquad \square$$

Аналогично доказывается следующее предложение.

Предложение 1.50. Для любых нечетных чисел $a_1, \dots, a_n$ выполнено сравнение:

$$(a_1^2-1)+\dots+(a_n^2-1) \equiv a_1^2 \dots a_n^2 - 1 \pmod{16}.$$

Предложение 1.51. Для любого нечетного простого q и нечетного m выполнено равенство:

$$\left(\frac{q}{m}\right) = \left(\frac{m}{q}\right)(-1)^{\frac{m-1}{2}\cdot\frac{q-1}{2}}.$$

Доказательство. Пусть $m = p_1 \dots p_n$ — разложение на простые множители, среди которых могут быть одинаковые. Учитывая теорему 1.56, имеем:

$$\left(\frac{q}{m}\right) = \left(\frac{q}{p_1}\right) \dots \left(\frac{q}{p_n}\right) =$$

$$= \left(\frac{p_1}{q}\right) \dots \left(\frac{p_n}{q}\right)(-1)^{\left(\frac{p_1-1}{2}+\dots+\frac{p_n-1}{2}\right)\frac{q-1}{2}}.$$

Так как (предложение 1.49):

$$\frac{m-1}{2} = \frac{p_1 \dots p_n - 1}{2} = \frac{p_1-1}{2} + \dots + \frac{p_n-1}{2} + 2t, \quad t \in \mathbb{Z},$$

то

$$\left(\frac{q}{m}\right) = \left(\frac{m}{q}\right)(-1)^{\frac{m-1}{2}\cdot\frac{q-1}{2}}. \qquad \square$$

Теорема 1.57 (свойства символа Якоби).

1. Если $a \equiv b \pmod m$, то $\left(\frac{a}{m}\right) = \left(\frac{b}{m}\right)$.
2. $\left(\frac{a^2}{m}\right) = 1$, в частности $\left(\frac{1}{m}\right) = 1$.
3. $\left(\frac{-1}{m}\right) = (-1)^{\frac{m-1}{2}}$.

4. $\left(\dfrac{2}{m}\right) = (-1)^{\frac{m^2-1}{8}}$.

5. $\left(\dfrac{a_1 \dots a_n}{m}\right) = \left(\dfrac{a_1}{m}\right) \dots \left(\dfrac{a_n}{m}\right)$.

6. Если m и a нечетны и больше единицы, то:

$$\left(\frac{a}{m}\right) = \left(\frac{m}{a}\right) (-1)^{\frac{m-1}{2} \cdot \frac{a-1}{2}}.$$

Доказательство.

1. Следует из теоремы 1.54.

2. Очевидно.

3. Пусть $m = p_1 \dots p_n$ — разложение на простые множители, среди которых могут быть одинаковые. Учитывая теорему 1.54 и равенства (предложение 1.49):

$$\frac{m-1}{2} = \frac{p_1 \dots p_n - 1}{2} = \frac{p_1 - 1}{2} + \dots + \frac{p_n - 1}{2} + 2t, \quad t \in \mathbb{Z},$$

получаем:

$$\left(\frac{-1}{m}\right) = \left(\frac{-1}{p_1}\right) \dots \left(\frac{-1}{p_n}\right) = (-1)^{\frac{p_1-1}{2} + \dots + \frac{p_n-1}{2}} = (-1)^{\frac{m-1}{2}}.$$

4. Доказывается аналогично, как и пункт 3, с использованием следствия 1.15 и предложения 1.50.

5. Следует из теоремы 1.54.

6. Пусть $a = q_1 \dots q_s$, где q_i — простые числа. Тогда, учитывая предложения 1.49 и 1.51, получаем:

$$\left(\frac{a}{m}\right) = \left(\frac{q_1 \dots q_s}{m}\right) = \left(\frac{q_1}{m}\right) \dots \left(\frac{q_s}{m}\right) =$$

$$= \left(\frac{m}{q_1}\right) \dots \left(\frac{m}{q_s}\right) (-1)^{\left(\frac{q_1-1}{2} + \dots + \frac{q_s-1}{2}\right)\frac{m-1}{2}} = \left(\frac{m}{a}\right) (-1)^{\frac{a-1}{2} \cdot \frac{m-1}{2}}. \quad \square$$

Замечание 1.6. Из теоремы 1.40 следует, что если $\left(\dfrac{a}{m}\right) = 1$ для символа Якоби, то это не означает, что квадратичное сравнение $x^2 - a \equiv 0 \pmod{m}$ имеет решение.

Символ Якоби является вспомогательным средством для вычисления символа Лежандра и позволяет во многих случаях упростить его вычисление, не применяя факторизацию числа, а лишь выделяя в числе степень двойки.

Алгоритм 1.5 (эффективное вычисление символа Лежандра на основе символа Якоби).

Вход: $a \in \mathbb{Z}$, m — целое нечетное число, $(a, m) = 1$.

Выход: $\left(\frac{a}{m}\right)$.

1. Заменить a на такое b, что $a \equiv b \pmod m$ и $|b| < \frac{m}{2}$.

2. Найти представление числа b в виде: $b = (-1)^s 2^t c$, $(2, c) = 1$. Тогда:

$$\left(\frac{a}{m}\right) = \left(\frac{b}{m}\right) = \left(\frac{-1}{m}\right)^s \left(\frac{2}{m}\right)^t \left(\frac{c}{m}\right).$$

3. Вычислить $\left(\frac{-1}{m}\right)$, $\left(\frac{2}{m}\right)$, используя теорему 1.57.

4. Символ $\left(\frac{c}{m}\right)$ выразить через $\left(\frac{m}{c}\right)$, используя теорему 1.57. Далее для вычисления $\left(\frac{m}{c}\right)$ снова перейти к шагу 1.

1.22. Вычисление квадратного корня

В предыдущем параграфе рассматривался вопрос разрешимости сравнения (1.55). Теперь покажем, как найти решение данного сравнения.

Для начала заметим, что если $p \equiv 3 \pmod 8$ или $p \equiv 7 \pmod 8$, то $p \equiv 3 \pmod 4$. Также если $p \equiv 1 \pmod 8$ или $p \equiv 5 \pmod 8$, то $p \equiv 1 \pmod 4$.

Далее нам понадобится следующее вспомогательное утверждение.

Предложение 1.52. Пусть p — нечетное простое число, a — целое, $p \nmid a$ и сравнение $x^2 \equiv a \pmod{p}$ имеет решение x. Тогда выполнены следующие утверждения:

1) если $p \equiv 3 \pmod{4}$, то $x \equiv \pm a^{\frac{p+1}{4}} \pmod{p}$;

2) если $p \equiv 5 \pmod{8}$, то

2.1) если $a^{\frac{p-1}{4}} \equiv 1 \pmod{p}$, то $x \equiv \pm a^{\frac{p+3}{8}} \pmod{p}$;

2.2) если $a^{\frac{p-1}{4}} \equiv -1 \pmod{p}$, то $x \equiv \pm 2a(4a)^{\frac{p-5}{8}} \pmod{p}$.

Доказательство. 1. Пусть $p \equiv 3 \pmod{4}$ и число x удовлетворяет сравнению $x \equiv a^{\frac{p+1}{4}} \pmod{p}$. Тогда, учитывая критерий Эйлера (теорема 1.51), получаем:

$$x^2 \equiv a^{\frac{p+1}{2}} = a \cdot a^{\frac{p-1}{2}} \equiv a \cdot \left(\frac{a}{p}\right) = a \pmod{p}.$$

2. Напомним, что из $p \equiv 5 \pmod{8}$ следует $p \equiv 1 \pmod{4}$. Учитывая критерий Эйлера, получаем:

$$0 \equiv a^{\frac{p-1}{2}} - 1 = (a^{\frac{p-1}{4}} - 1)(a^{\frac{p-1}{4}} + 1) \pmod{p}.$$

Поэтому либо $a^{\frac{p-1}{4}} \equiv 1 \pmod{p}$, либо $a^{\frac{p-1}{4}} \equiv -1 \pmod{p}$.

2.1. Пусть $a^{\frac{p-1}{4}} \equiv 1 \pmod{p}$ и число x удовлетворяет сравнению $x \equiv a^{\frac{p+3}{8}} \pmod{p}$. Тогда:

$$x^2 \equiv a^{\frac{p+3}{4}} = a \cdot a^{\frac{p-1}{4}} \equiv a \pmod{p}.$$

2.2. Пусть $a^{\frac{p-1}{4}} \equiv -1 \pmod{p}$ и $x \equiv 2a(4a)^{\frac{p-5}{8}} \pmod{p}$. Тогда:

$$x^2 \equiv 4a^2 \cdot (4a)^{\frac{p-5}{4}} = a \cdot (4a)^{1+\frac{p-5}{4}} = a \cdot 2^{\frac{p-1}{2}} \cdot a^{\frac{p-1}{4}} \pmod{p}.$$

Из теоремы 1.51 и следствия 1.15 следует, что:

$$2^{\frac{p-1}{2}} = \left(\frac{2}{p}\right) = -1.$$

Поэтому:

$$x^2 \equiv a \cdot 2^{\frac{p-1}{2}} \cdot a^{\frac{p-1}{4}} = a(-1)(-1) = a \pmod{p}. \qquad \square$$

Из предложения 1.52 следует, что нерассмотренным остался случай $p \equiv 1 \pmod{8}$, для которого неизвестен простой способ определения решения сравнения (1.55).

Предложение 1.53. Пусть $p = 2^n q + 1$ — простое число, q — нечетное целое, $n \geqslant 1$ и a — некоторое целое число, для которого $(a, p) = 1$. Тогда:

– либо $a^q \equiv 1 \pmod{p}$;

– либо найдется такое целое число k, $0 \leqslant k < n$, для которого $a^{2^k q} \equiv -1 \pmod{p}$.

Доказательство. Учитывая малую теорему Ферма, получаем:

$$0 \equiv a^{p-1} - 1 = a^{2^n q} - 1 = \left(a^{2^{n-1} q} - 1\right)\left(a^{2^{n-1} q} + 1\right) =$$

$$= \left(a^{2^{n-2} q} - 1\right)\left(a^{2^{n-2} q} + 1\right)\left(a^{2^{n-1} q} + 1\right) =$$

$$\ldots$$

$$= (a^q - 1)(a^q + 1)\left(a^{2q} + 1\right) \ldots \left(a^{2^{n-1} q} + 1\right) \pmod{p}.$$

Осталось заметить, что одна из данных скобок делится на p (предложение 1.12). □

Вернемся к сравнению (1.55). Пусть $p \equiv 1 \pmod 8$. Если выполнено первое утверждение предложения 1.53 ($a^q \equiv 1 \pmod p$), то решение сравнения (1.55) определяется так:

$$x \equiv \pm a^{\frac{q+1}{2}} \pmod{p}.$$

Действительно:

$$x^2 \equiv a \cdot a^q \equiv a \pmod{p}.$$

Остался случай $a^{2^k q} \equiv -1 \pmod{p}$, $0 \leqslant k < n$. В данном случае необходимо найти такое число w, для которого выполнено $w^2 a^q \equiv 1 \pmod p$. В этом случае решение сравнения (1.55) будет иметь такой вид:

$$x \equiv \pm w a^{\frac{q+1}{2}} \pmod{p},$$

так как:

$$x^2 \equiv a \cdot w^2 \cdot a^q \equiv a \pmod{p}.$$

Нам понадобится следующее утверждение.

Лемма 1.11. Пусть $p = 2^n q + 1$ — простое число, q — нечетное целое, $n \geqslant 1$ и b — произвольный квадратичный невычет по модулю p. Тогда верны следующие утверждения.

1. Обозначим $z \equiv b^q \pmod{p}$. Тогда $P_p(z) = 2^n$ — показатель числа z по модулю p.

2. Обозначим $z_k \equiv z^{2^k} \pmod{p}$. Тогда $z_k \equiv z_{k-1}^2 \pmod{p}$ и $P_p(z_k) = 2^{n-k}$.

3. Пусть u, v — такие целые числа, что $P_p(u) = P_p(v) = 2^{k+1}$. Тогда $P_p(uv) \,|\, 2^k$.

Доказательство. 1. Учитывая малую теорему Ферма, получаем:

$$z^{2^n} \equiv b^{p-1} \equiv 1 \pmod{p}.$$

Поэтому из теоремы 1.41 следует, что $P_p(z)$ является делителем числа 2^n. При этом

$$z^{2^{n-1}} \equiv b^{\frac{p-1}{2}} \equiv \left(\frac{b}{p}\right) = -1 \pmod{p}.$$

Следовательно, $P_p(z) = 2^n$.

2. Понятно, что $z_{k-1}^2 \equiv \left(z^{2^{k-1}}\right)^2 = z^{2^k} \equiv z_k \pmod{p}$. При этом из первого пункта видно, что $P_p(z_k) = \dfrac{P_p(z)}{2^k} = 2^{n-k}$.

3. Поскольку $P_p(u) = 2^{k+1}$, то из сравнений:

$$0 \equiv u^{2^{k+1}} - 1 = \left(u^{2^k} - 1\right)\left(u^{2^k} + 1\right) \pmod{p}$$

следует, что $u^{2^k} \equiv -1 \pmod{p}$. Аналогично, $v^{2^k} \equiv -1 \pmod{p}$. Перемножая указанные сравнения, получаем $(uv)^{2^k} \equiv 1 \pmod{p}$. Осталось воспользоваться теоремой 1.41. $\square$

Итак, пусть для некоторого целого k, $0 \leqslant k < n$, выполнено сравнение $a^{2^k q} \equiv -1 \pmod{p}$. Заметим, что показатель элемента a^q по модулю p равен 2^{k+1}. Действительно, из последнего сравнения следует, что $a^{2^{k+1} q} \equiv 1 \pmod{p}$, поэтому $P(a^q) \,|\, 2^{k+1}$, при этом $\left(a^q\right)^{2^k} \equiv -1 \pmod{p}$. Поэтому $P(a^q) = 2^{k+1}$.

Определим конечную последовательность $u_1,\ldots,u_{r+1}$ сравнениями:

$$\begin{aligned} &u_1 \equiv a^q \pmod p,\\ &u_2 \equiv z_{i_1}u_1 \pmod p,\\ &\ldots\\ &u_{r+1} \equiv z_{i_r}u_r \pmod p, \end{aligned}$$

где индексы i_j выбираются из условия $P(u_j) = P(z_{i_j})$, $i = 1,\ldots,r$, а числа z_{i_j} определены в лемме 1.11. Из пункта 2 леммы 1.11 следует, что:

$$P(u_j) \mid P(u_{j-1}), \quad j = 2,\ldots,r,$$

и выполнена цепочка неравенств:

$$P(u_1) > P(u_2) > \ldots > P(u_{r+1}) = 1$$

для некоторого $r \geqslant 1$. Таким образом, получаем:

$$1 \equiv u_{r+1} \equiv z_{i_1}\ldots z_{i_r}a^q \pmod p.$$

Так как $z_{i_j} \equiv z_{i_{j-1}}^2 \pmod p$ (лемма 1.11), то, обозначив:

$$w = z_{i_1-1}\ldots z_{i_r-1},$$

получаем:

$$1 \equiv u_{r+1} \equiv (z_{i_1-1}\ldots z_{i_r-1})^2 a^q \equiv w^2 a^q \pmod p.$$

Из этого следует, что решением сравнения (1.55) будет:

$$x \equiv \pm w a^{\frac{q+1}{2}} \pmod p.$$

Обобщая все изложенное выше, приведем алгоритм, который позволяет находить решения сравнения (1.55).

Алгоритм 1.6 (алгоритм Тонелли-Шенкса).
Вход: $a \in \mathbb{Z}$, нечетное простое $p = 2^n q+1$, q — нечетное целое, $\left(\frac{a}{p}\right) = 1$.
Выход: такое целое x, что $x^2 \equiv a \pmod p$.
1. Если $p \equiv 3 \pmod 4$, то $x \equiv a^{\frac{p+1}{4}} \pmod p$ и остановиться.

2. Если $p \equiv 5 \pmod 8$, то

2.1. если $a^{\frac{p-1}{4}} \equiv 1 \pmod p$, то $x \equiv a^{\frac{p+3}{8}} \pmod p$ и остановиться;

2.2. если $a^{\frac{p-1}{4}} \equiv -1 \pmod p$, то $x \equiv 2a(4a)^{\frac{p-5}{8}} \pmod p$ и остановиться.

3. Выбирая случайным образом, найти квадратичный невычет b и определить:

$$z \equiv b^q \pmod p, \quad r = n.$$

4. Определить:

$$x \equiv a^{\frac{q+1}{2}} \pmod p, \quad u \equiv a^q \pmod p.$$

5. Цикл: пока $u \not\equiv 1 \pmod p$

5.1. Найти наименьшее m, при котором $u^{2^m} \equiv 1 \pmod p$.

5.2. Определить:

$$z_{pred} \equiv z^{2^{r-m-1}} \pmod p, \quad z_{tek} \equiv z_{pred}^2 \pmod p,$$

$$x \equiv xz_{pred} \pmod p, \quad u \equiv uz_{tek} \pmod p, \quad r = m.$$

Заметим, что в данном алгоритме $P(u) = P(z_{tek}) = 2^m$. После нахождения решения x, второе решение, очевидно, $p - x$.

1.23. Тесты на простоту

Алгоритмы строго доказательства простоты заданного нечетного числа n являются трудоемкими. Такие алгоритмы называются *детерминированными*. Гораздо эффективнее могут быть реализованы алгоритмы, которые проверяют, не является ли некоторое число n составным с некоторой вероятностью — *вероятностные тесты*.

1.23.1. Тест на основе малой теоремы Ферма

Самой очевидной идеей построения такого теста является применение малой теоремы Ферма. Согласно этой теореме простота числа n влечет выполнение сравнения:

$$a^{n-1} \equiv 1 \pmod n, \tag{1.65}$$

для всех целых чисел a, для которых $(a, n) = 1$. Поэтому если найдется целое a, взаимно простое с n и такое, что выполнено $a^{n-1} \not\equiv 1 \pmod{n}$, то n — составное.

Определение 1.30. Пусть $a, n \in \mathbb{N}$, n — нечетное составное, $(a, n) = 1$. Число n называется *псевдопростым* по основанию a, если для a и n выполнено сравнение (1.65).

Например, число $n = 21$ является псевдопростым по модулю $a = 13$.

Предложение 1.54. Пусть n — нечетное составное натуральное число. Если для некоторого a, $(a, n) = 1$, число n не является псевдопростым по основанию a, то n не является псевдопростым относительно, по крайней мере, половины вычетов из приведенной системы вычетов по основанию n.

Доказательство. Пусть b — некоторое целое число, взаимно простое с n, по основанию которого число n является псевдопростым, a — некоторое целое число, $(a, n) = 1$, по основанию которого число n не является псевдопростым. Тогда n не является псевдопростым по основанию ab, так как:

$$(ab)^{n-1} = a^{n-1}b^{n-1} \equiv a^{n-1} \not\equiv 1 \pmod{n}.$$

Пусть $b_1, \ldots, b_k$ — некоторые (попарно несравнимые) числа из приведенной системы вычетов по модулю n, по основанию которых число n является псевдопростым. Тогда, учитывая лемму 1.8, $ab_1, \ldots, ab_k$ — (попарно несравнимые) числа из приведенной системы вычетов по модулю n, по основанию которых число n не является псевдопростым. □

В 1885 году немецкий математик Альвин Корсельт привел критерий составных чисел n, для которых малая теорема Ферма выполнена для всех целых чисел a, взаимно простых с n. Но он не предъявил в явном виде ни одного такого числа. Впервые это сделал Роберт Кармайкл. В настоящее время числа, удовлетворяющие критерию Корсельта, называются числами Кармайкла.

Определение 1.31. Нечетное составное число n называется *числом Кармайкла*, если сравнение (1.65) выполнено для любого a такого, что $(a, n) = 1$.

Заметим, что числа Кармайкла n являются псевдопростыми по любому основанию a, $(a, n) = 1$.

Теорема 1.58 (критерий Корсельта). Нечетное составное число n является числом Кармайкла тогда и только тогда, когда выполнены следующие условия:

(i) число n свободно от квадратов, т.е. для любого простого делителя p числа n выполнено $p^2 \nmid n$;

(ii) если n представимо в виде $n = p_1 \dots p_k$, то для всех $i = 1, \dots, k$ выполнено условие $(p_i - 1) \mid (n - 1)$.

Доказательство. *Достаточность.* Пусть выполнены условия (i), (ii) и a — произвольное целое число, $(a, n) = 1$. Возведя сравнение $a^{p_i - 1} \equiv 1 \pmod{p_i}$ в степень $\dfrac{n-1}{p_i - 1}$, получим $a^{n-1} \equiv 1 \pmod{p_i}$, $i = 1, \dots, k$. Применяя теорему 1.26, получим $a^{n-1} \equiv 1 \pmod{[p_1, \dots, p_k] = n}$.

Необходимость. Пусть нечетное составное число $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ является числом Кармайкла. Пусть a_i — первообразный корень по модулю p_i, $i = 1, \dots, k$ (теорема 1.49). Тогда (учитывая теорему Дирихле 1.20) из сравнений:

$$a_i^{n-1} \equiv 1 \pmod{n} \equiv 1 \pmod{p_i}$$

следует, что $p_i - 1 = P(a_i) \mid n - 1$ (теорема 1.41). Поэтому условие (ii) выполнено.

Предположим, что $\alpha_j > 1$ для некоторого индекса j. Пусть b_j — первообразный корень по модулю $p_j^{\alpha_j}$ (теорема 1.49). Тогда:

$$p_j^{\alpha_j - 1}(p_j - 1) = P(b_j) \mid n - 1, \quad p_j \mid n - 1.$$

При этом $p_j - 1 \mid n - 1$, что невозможно в силу неравенства $p_j > 2$. Противоречие. Поэтому условие (i) тоже выполнено. $\square$

Следствие 1.16. Пусть нечетное число $n = p_1 \dots p_k$ является числом Кармайкла. Тогда $k \geqslant 3$.

Доказательство. Предположим, что $n = pq$ — число Кармайкла, $2 < p < q$. При этом $p-1 \,|\, n-1$, $q-1 \,|\, n-1$. Запишем число $n-1$ в виде $n-1 = pq+p-p-1 = p(q-1)+p-1$. Получаем $n-1 \equiv p-1 \pmod{q-1}$. Из этого и $q-1 \,|\, n-1$ следует, что $q-1 \,|\, p-1$, что невозможно в силу того, что $2 < p < q$. Противоречие. □

Наименьшим числом Кармайкла является $561 = 3 \cdot 11 \cdot 17$. Известно, что чисел Кармайкла бесконечно много [48], но встречаются они достаточно редко.

Пусть нечетное составное число n не является числом Кармайкла. Тогда вероятность того, что составное число n пройдет тест (1.65) для случайно выбранного a, $(a, n) = 1$, не превышает $1/2$. Можно рассмотреть следующий вероятностный метод проверки числа на простоту. Выбираем случайное число в интервале $1 < a < n$ и вычисляем (a, n) по алгоритму Евклида. Если $(a, n) > 1$, то n простым не является. В противном случае с помощью алгоритма быстрого возведения в степень вычисляем $a^{n-1} \pmod{n}$ и сравниваем с единицей. Если сравнение не выполнено, то число n — составное. Если же тест (1.65) проходит, то повторяем процедуру с другим значением a. Вероятность того, что составное число n пройдет тест k раз не превышает $1/2^k$.

1.23.2. Тест Соловея-Штрассена

В 1977 году Соловеем и Штрассеном был опубликован тест, основанный на свойствах символов Лежандра и Якоби. Следующий результат лежит в основе этого теста.

Теорема 1.59 (**критерий Эйлера**)**.** Натуральное нечетное $n > 1$ является простым тогда и только тогда, когда для любого целого a, $(a, n) = 1$, выполнено сравнение:

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}, \tag{1.66}$$

где $\left(\frac{a}{n}\right)$ — символ Якоби.

Доказательство. *Необходимость* следует из теоремы 1.54.

Достаточность. Покажем, что если n — нечетное составное число, то найдется хотя бы одно целое число a, $(a, n) = 1$, для которого сравнение (1.66) не выполнено. Пусть $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ — каноническое разложение числа n. Возможны следующие случаи.

1. Найдется такой индекс j, $1 \leqslant j \leqslant k$, для которого $\alpha_j > 1$. Без ограничения общности можно считать, что $j = 1$.

Рассмотрим число $a = 1 + \dfrac{n}{p_1}$. Понятно, что $(a, n) = 1$, так как при делении a на p_i получаем в остатке единицу: $a \equiv 1 \pmod{p_i}$, $i = 1, \dots, k$. Значит a — квадратичный вычет по модулю p_i, так как корнем уравнения $x^2 \equiv a \pmod{p_i}$ является $x \equiv 1 \pmod{p_i}$, $i = 1, \dots, k$. Поэтому:

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{\alpha_1} \dots \left(\frac{a}{p_k}\right)^{\alpha_k} = 1.$$

Покажем, что $a^{\frac{n-1}{2}} \not\equiv 1 \pmod{n}$. Предположим противное: пусть $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$. Из последнего сравнения следует, что выполнено $P(a) \mid \dfrac{n-1}{2}$ (теорема 1.41, $P(a)$ — показатель числа a по модулю n). Следовательно, $P(a) \mid n - 1$. Покажем, что $p_1 \mid P(a)$. Действительно, так как $a = 1 + p_1^{\alpha_1 - 1} t$, $t \in \mathbb{Z}$, $(t, p_1) = 1$, то:

$$1 \equiv a^{P(a)} = \left(1 + p_1^{\alpha_1 - 1} t\right)^{P(a)} \equiv 1 + p_1^{\alpha_1 - 1} t P(a) \pmod{p_1^{\alpha_1}}.$$

Поэтому $p_1^{\alpha_1 - 1} t P(a) \equiv 0 \pmod{p_1^{\alpha_1}}$ или, что равносильно, $p_1 \mid t P(a)$. Так как $(t, p_1) = 1$, то $p_1 \mid P(a) \mid n - 1$. При этом $p_1 \mid n$. Противоречие, так как $p_1 > 2$.

2. $\alpha_1 = \dots = \alpha_k = 1$, $n = p_1 \dots p_k$. Определим целое число a как решение системы сравнений:

$$\begin{cases} x \equiv b \pmod{p_1}, \\ x \equiv 1 \pmod{p_2}, \\ \dots \\ x \equiv 1 \pmod{p_k}, \end{cases}$$

где b — произвольный квадратичный невычет по модулю p_1. Получаем, что $\left(\frac{a}{p_1}\right) = \left(\frac{b}{p_1}\right) = -1$, $\left(\frac{a}{p_i}\right) = 1$, $i = 2, \ldots, k$:

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\ldots\left(\frac{a}{p_k}\right) = -1.$$

С другой стороны, $a \equiv 1 \pmod{p_i}$, $i = 2, \ldots, k$, поэтому $a^{\frac{n-1}{2}} \equiv 1 \pmod{p_i}$, $i = 2, \ldots, k$. Следовательно, $a^{\frac{n-1}{2}} \equiv 1 \pmod{p_2 \ldots p_k}$ (китайская теорема об остатках).

Покажем, что $a^{\frac{n-1}{2}} \not\equiv -1 \pmod{n}$. Предположим, что это не так: $a^{\frac{n-1}{2}} \equiv -1 \pmod{n}$. Тогда $a^{\frac{n-1}{2}} \equiv -1 \pmod{p_2 \ldots p_k}$, так как $p_2 \ldots p_k \mid n$. Противоречие. □

Определение 1.32. Нечетное составное число n называется *эйлеровым псевдопростым по основанию* a, если для чисел a и n, $(a, n) = 1$, выполнено сравнение (1.66).

Замечание 1.7. 1. Очевидно, что если n — эйлерово псевдопростое число по основанию a, то n — псевдопростое по основанию a.

2. В силу теоремы 1.59 аналогов чисел Кармайкла (т.е. нечетных составных чисел, которые были бы псевдопростыми по любому основанию) не существует.

Предложение 1.55. Пусть n — нечетное составное натуральное число. Тогда n не является эйлеровым псевдопростым относительно, по крайней мере, половины вычетов из приведенной системы вычетов по модулю n.

Доказательство. Пусть b — некоторое целое число, взаимно простое с n, по основанию которого число n является эйлеровым псевдопростым, a — некоторое целое число, $(a, n) = 1$, по основанию которого число n не является эйлеровым псевдопростым (такое число существует, см. теорему 1.59). Тогда n не является эйлеровым псевдопростым по основанию ab. Действительно, так как:

$$(ab)^{\frac{n-1}{2}} = a^{\frac{n-1}{2}} b^{\frac{n-1}{2}},$$

$$\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right) \equiv \left(\frac{a}{n}\right) b^{\frac{n-1}{2}} \pmod{n},$$

то $(ab)^{\frac{n-1}{2}} \not\equiv \left(\frac{ab}{n}\right) \pmod{n}$.

Пусть $b_1, \ldots, b_k$ — некоторые (попарно несравнимые) числа из приведенной системы вычетов по модулю n, по основанию которых число n является эйлеровым псевдопростым. Тогда, учитывая лемму 1.8, $ab_1, \ldots, ab_k$ — (попарно несравнимые) числа из приведенной системы вычетов по модулю n, по основанию которых число n не является эйлеровым псевдопростым. □

Следующий тест основан на теореме 1.59 и предложении 1.55.

Алгоритм 1.7 (**тест Соловея-Штрассена**).

Вход: натуральное нечетное число n.

Выход: заключение о том, что число составное, либо заключение о том, что число n не является составным с некоторой вероятностью.

Цикл $i = 1, \ldots, N$.

1. Сгенерировать случайным образом число a, $0 < a < n$.

2. Если $(a, n) > 1$, то завершить цикл и алгоритм с уведомлением, что число составное.

3. Если $a^{\frac{n-1}{2}} \not\equiv \left(\frac{a}{n}\right) \pmod{n}$, то завершить цикл и алгоритм с уведомлением, что число составное.

В данном алгоритме число итераций N определяет вероятность принять составное число n за простое. Данная вероятность не превосходит числа $\frac{1}{2^N}$. Следовательно, если алгоритм завершит работу с заключением, что число n (вероятно) простое, то оно является простым с вероятностью не менее чем $1 - \frac{1}{2^N}$. Для теста можно взять значение $N = 20$.

Чаще всего на практике для снижения трудоемкости алгоритма выбирают число a не из интервала $0 < a < n$, а из меньшего интервала $0 < a < c$, где константа c определяет максимально возможное число из стандартных целочисленных типов данных.

Сложность теста Соловея-Штрассена определяется сложностью вычисления символа Якоби и равна $O(N \log^3 n)$.

1.23.3. Тест Миллера-Рабина

Теорема 1.60 (Миллер, 1976). Пусть n — нечетное натуральное число и $n-1=2^m q$, $(2,q)=1$. Тогда следующие утверждения равносильны:

1) число n является простым;

2) для любого целого a такого, что $(a,n)=1$, либо $a^q \equiv 1 \pmod{n}$, либо найдется такое целое число k, $0 \leqslant k < m$, что выполнено $a^{2^k q} \equiv -1 \pmod{n}$.

Доказательство. Необходимое условие простоты следует из предложения 1.53.

Обратно, предположим, что условие 2 не выполнено. Тогда найдется такое целое a, $(a,n)=1$, для которого $a^q \not\equiv 1 \pmod{n}$ и $a^{2^k q} \not\equiv -1 \pmod{n}$ для любого целого k, $0 \leqslant k < m$. Тогда последовательность:

$$x_0 = a^q, \quad x_1 = x_0^2 = a^{2q}, \ldots, \; x_{m-1} = x_{m-2}^2 = a^{2^{m-1}q} \pmod{n}$$

не содержит -1. Определим значение:

$$x_m = x_{m-1}^2 = a^{n-1} \pmod{n}.$$

Для значения x_m возможны следующие случаи.

1. $x_m \not\equiv 1 \pmod{n}$. Тогда из малой теоремы Ферма следует, что n — непростое.

2. $x_m \equiv 1 \pmod{n}$. Тогда n не может быть простым, так как в случае из простоты n из сравнения $x_m \equiv 1 \pmod{n}$ следует, что либо выполнено $x_0 \equiv 1 \pmod{n}$, либо $x_k \equiv -1 \pmod{n}$ для некоторого k, $0 \leqslant k < m$. □

Пусть n — нечетное составное натуральное число и $n-1=2^m q$, $(2,q)=1$. Если для некоторого целого a, $(a,n)=1$, либо $a^q \equiv 1 \pmod{n}$, либо найдется такое целое число k, $0 \leqslant k < m$, что $a^{2^k q} \equiv -1 \pmod{n}$, то число n называется *сильно псевдопростым* по основанию a.

Предложение 1.56. Пусть n — нечетное составное натуральное число. Если n является сильно псевдопростым по основанию a, то ото также является эйлеровым псевдопростым по основанию a.

Доказательство данного утверждения можно найти, например, в [24].

Теорема 1.61 (Рабин, 1980). Пусть n — нечетное составное натуральное число, $(6, n) = 1$, $n - 1 = 2^m q$, $(2, q) = 1$, $S \subset \mathbb{Z}_n^*$ — множество всех таких элементов a, для которых выполнено хотя бы одно из условий:

1) $a^q \equiv 1 \pmod{n}$;

2) найдется такое целое число k, $0 \leqslant k < m$, что $a^{2^k q} \equiv -1 \pmod{n}$.

Тогда мощность множества S не превосходит $\dfrac{n}{4}$.

Доказательство. Приведем доказательство данного утверждения, следуя идеям работ [12, 29].

Определим множество $A \subset \mathbb{Z}_n^*$ всех таких элементов a, для которых выполнено одно из двух условий:

1. $a^{n-1} \not\equiv 1 \pmod{n}$.

2. Число a является первообразным корнем по модулю p — некоторого простого делителя числа n и выполнено условие $a^z \not\equiv -1 \pmod{n}$ для любого целого z.

Лемма 1.12. Для любых $a \in A$, $s \in S$ выполнено условие $as \pmod{n} \notin S$, что равносильно, $aS \cap S = \varnothing$.

Доказательство. Заметим, что для любого $s \in S$ выполнено $s^{n-1} \equiv 1 \pmod{n}$, так как $n - 1 = 2^s q$. Рассмотрим два случая относительно элемента $a \in A$.

1. Для a выполнено первое условие: $a^{n-1} \not\equiv 1 \pmod{n}$. Тогда:

$$(as)^{n-1} \equiv a^{n-1} s^{n-1} \equiv a^{n-1} \not\equiv 1 \pmod{n}.$$

Поэтому $as \notin S$, так как не выполнено необходимое условие принадлежности множеству S.

2. Пусть для a выполнено второе условие: $a^{n-1} \equiv 1 \pmod{n}$, $a^z \not\equiv -1 \pmod{n}$ для любого целого z, причем a — первообразный корень по некоторому простому делителю p числа n.

Докажем от противного. Предположим, что $as \in S$. Тогда относительно s и as возможны следующие варианты.

2.1. $s^q \equiv 1 \pmod{n}$ и $(as)^q \equiv 1 \pmod{n}$. Тогда $a^q \equiv 1 \pmod{n}$, так как:

$$1 \equiv (as)^q \equiv a^q s^q \equiv a^q \pmod{n}.$$

Поэтому $a^q \equiv 1 \pmod{p}$ и $(p-1) \mid q$, так как a — первообразный корень по модулю p. Но $(p-1) \nmid q$, так как $p-1$ четно, а q нечетно.

2.2. $s^q \equiv 1 \pmod{n}$ и $(as)^{2^l q} \equiv -1 \pmod{n}$ для некоторого $0 \leqslant l < m$. Тогда:

$$-1 \equiv (as)^{2^l q} \equiv a^{2^l q}(s^q)^{2^l} \equiv a^{2^l q} \pmod{n},$$

что невозможно в силу выбора a.

2.3. $s^{2^k q} \equiv -1 \pmod{n}$ для некоторого $0 \leqslant k < m$ и $(as)^q \equiv 1 \pmod{n}$. Из второго сравнения выразим $s^q \equiv a^{-q} \pmod{n}$ и подставим полученное значение в первое сравнение: $a^{-2^k q} \equiv -1 \pmod{n}$, что противоречит выбору a.

2.4. $s^{2^k q} \equiv -1 \pmod{n}$ для некоторого $0 \leqslant k < m$ и $(as)^{2^l q} \equiv -1 \pmod{n}$ для некоторого $0 \leqslant l < m$. Если $k < l$, то:

$$-1 \equiv (as)^{2^l q} \equiv a^{2^l q} \left(s^{2^k q}\right)^{2^{l-k}} \equiv a^{2^l q} \pmod{n},$$

что противоречит выбору a. Если $k \geqslant l$, то выразим $s^{2^l q} \equiv -a^{-2^l q} \pmod{n}$ и подставим это выражение в первое сравнение:

$$-1 \equiv s^{2^k q} \equiv \left(s^{2^l q}\right)^{2^{k-l}} \equiv \left(-a^{-2^l q}\right)^{2^{k-l}} \pmod{n}.$$

При $k > l$ получаем $a^{-2^k q} \equiv -1 \pmod{n}$, что противоречит выбору a. Если же $k = l$, то $a^{2^k q} \equiv 1 \pmod{n}$. Так как показатель числа a по модулю p равен $p-1$, то $(p-1) \mid 2^k q$. Поэтому $2^k q = (p-1)t$ для некоторого целого t и:

$$s^{2^k q} \equiv (s^{p-1})^t \equiv 1 \pmod{p},$$

что противоречит выбору s. □

Лемма 1.13. Пусть $a, b \in \mathbb{Z}_n^*$, $a \neq b$, $b \in A$. Если $a^{-1}b \in A$, то $aS \cap bS = \varnothing$.

Доказательство. Предположим противное: $aS \cap bS \neq \varnothing$. Тогда для некоторых $s_1, s_2 \in S$ выполнено сравнение $as_1 \equiv bs_2 \pmod{n}$, из которого следует, что $s_1 \equiv a^{-1}bs_2 \pmod{n}$. Так как $a^{-1}b \in A$, то получили противоречие с леммой 1.12. $\square$

Продолжение доказательства теоремы 1.61. Относительно числа n возможны следующие случаи.

1. Найдется такое простое число p, для которого $p^\alpha \,|\, n$, $\alpha > 1$. Рассмотрим множество:

$$G = \left\{ 1 + k\frac{n}{p} \ (mod\ n) \mid 0 < k < p \right\}.$$

Множество G является подгруппой мультипликативной группы кольца вычетов $\mathbb{Z}_n^*$. Действительно, из сравнений:

$$\left(1 + k\frac{n}{p}\right)\left(1 + t\frac{n}{p}\right) \equiv \left(1 + (k+t)\frac{n}{p} + kt\frac{n^2}{p^2}\right) \equiv$$

$$\equiv \left(1 + m\frac{n}{p}\right) \pmod{n},$$

где $m \equiv k + t \pmod{p}$, следует, что G замкнуто относительно операции умножения. Поэтому по критерию конечной подгруппы G — мультипликативная группа (предложение 2.16).

Пусть $g = \left(1 + k\frac{n}{p}\right)$ — отличный от единицы элемент группы G, $0 < k < p$. Тогда:

$$g^{n-1} = \left(1 + k\frac{n}{p}\right)^{n-1} \equiv 1 + k(n-1)\frac{n}{p} \pmod{n}.$$

Так как $p \,|\, n$ и $p \nmid (n-1)$, то $n \nmid k(n-1)\frac{n}{p}$. Поэтому $g^{n-1} \not\equiv 1 \pmod{n}$. Это означает, что любой отличный от единицы элемент группы G принадлежит множеству A: $G \setminus \{1\} \subseteq A$. Тогда для любых $a, b \in G$, $a \neq b$, выполнено $a^{-1}b \in A$ и $aS \cap bS = \varnothing$ (лемма 1.13).

Таким образом:

$$|\mathbb{Z}_n| \geqslant \left|\bigcup_{g\in G} gS\right| = |G|\cdot|S| = p|S|, \quad |S| \leqslant \frac{n}{p}.$$

Так как $(n, 6) = 1$, то $p \geqslant 5$. Поэтому $|S| < n/4$.

2. Число n не делится на квадрат никакого простого числа: $n = p_1 \dots p_k$, $p_1 < \dots < p_k$, $k \geqslant 2$. Пусть a_i — первообразный корень по модулю p_i, $i = 1, \dots, k$ (теорема 1.44), c_i — решение системы сравнений (см. теорему 1.33):

$$\begin{cases} x \equiv 1 \pmod{p_1}, \\ \dots \\ x \equiv a_i \pmod{p_i}, \\ \dots \\ x \equiv 1 \pmod{p_k}, \end{cases}$$

где все правые части сравнений равны единицы, кроме сравнения с номером i. Заметим, что для любого целого z выполнено $c_i^z \not\equiv -1 \pmod{n}$. Действительно, если для некоторого целого z выполнено $c_i^z \equiv -1 \pmod{n}$, то $c_i^z \equiv -1 \pmod{p_j}$ для любого $j = 1, \dots, k$. Но для случаев $j \neq i$ это невозможно, так как $c_i \equiv 1 \pmod{p_j}$. Противоречие. Таким образом, $c_i, c_i^{-1} \in A$, $i = 1, \dots, k$.

Рассмотрим вычет $d \equiv c_1 c_2 \pmod{n}$. Покажем, что $d \in A$. Возможны случаи.

$k \geqslant 3$. Тогда для некоторого j выполнено:

$$c_1 \equiv 1 \pmod{p_j}, \quad c_2 \equiv 1 \pmod{p_j}.$$

Поэтому для любого целого z выполнено $d^z \not\equiv -1 \pmod{n}$, в противном бы случае, для некоторого целого z выполнено $d^z \equiv c_1^z c_2^z \equiv -1 \pmod{p_j}$, что невозможно.

$k = 2$. Покажем, что $d^{n-1} \not\equiv 1 \pmod{n}$. Предположим противное: пусть $d^{n-1} \equiv 1 \pmod{n}$. Тогда $d^{n-1} \equiv 1 \pmod{p_2}$. Так как $d \equiv a_2 \pmod{p_2}$, то d — первообразный корень по модулю p_2. Поэтому $(p_2 - 1) \mid (n - 1)$. Но из соотношений:

$$n - 1 = p_1 p_2 - 1 = p_1(p_2 - 1) + p_1 - 1, \quad 0 < p_1 - 1 < p_2 - 1,$$

следует, что $n-1$ не делится на p_2-1. Противоречие.

Таким образом, $d \in A$. Рассмотрим четыре множества S, c_1S, c_2S, dS. Данные множества не пересекаются (лемма 1.13). Следовательно, $4|S| \leqslant |\mathbb{Z}_n|$. $\square$

Следующий алгоритм является одним из лучших вероятностных тестов на простоту.

Алгоритм 1.8 (тест Миллера-Рабина).

Вход: натуральное нечетное число n, $n-1=2^m q$, $(2,q)=1$.

Выход: заключение о том, что число составное, либо заключение о том, что число не является составным с некоторой вероятностью.

Цикл $i=1,\ldots,N$.

1. Сгенерировать случайным образом число a, $0<a<n$.

2. Если $(a,n)>1$, то завершить цикл и алгоритм с уведомлением, что число составное.

3. Вычислить $x_0 = a^q \pmod n$. Если $x_0 \in \{1, n-1\}$, то перейти в шаг 1 (к следующему шагу итерации по переменной i).

4. Цикл $k=1,\ldots,m-1$.

Вычислить $x_k = x_{k-1}^2 \pmod n$. Если $x_k = n-1$, то перейти в шаг 1 (к следующему шагу итерации по переменной i).

Если $x_0 \notin \{1, n-1\}$ и $x_k \neq n-1$ для любого $k=1,\ldots,m-1$, то завершить алгоритм с уведомлением, что число составное.

Из теорем Миллера и Рабина следует, что мы можем принять составное число n за простое с вероятностью, не превышающей значения $1/4$. В алгоритме 1.8 реализовано N шагов итераций, следовательно, общая вероятность принять составное число за простое не превосходит значения $\dfrac{1}{4^N}$.

Сложность теста Миллера-Рабина равна $O(N \log^3 n)$.

Рассмотрим способ генерации простого числа, использующий вероятностные алгоритмы проверки чисел на простоту.

Алгоритм 1.9 (генерация простого числа).

Вход: k — разрядность искомого простого числа, N — количество итераций (проверок) для вероятностного теста на простоту.

Выход: число p с вероятностью не менее $1 - \frac{1}{4^N}$.

1. Сгенерировать случайное k-битное число $p = (b_{k-1} \dots b_1 b_0)_2$.

2. Положить $b_{k-1} = 1$, $b_0 = 1$.

3. Проверить, что число p не делится на 3,5,7,....

4. Проверить число p тестом Миллера-Рабина с параметром N. Если число p прошло тест, то вернуть p как результат. В противном случае вернуться в шаг 1.

Равенство $b_{k-1} = 1$ на втором шаге гарантирует, что длина числа p равна в точности k бит. Равенство $b_0 = 1$ необходимо, чтобы число p было нечетным.

Число проверок на шаге 3. варьируется от 256 до 2000. Например, если число p проверять на делимость на 3,5,7, то отбраковываются 54% нечетных составных чисел; при проверке в качестве делителей простых чисел, не превосходящих 100, — 76%; при проверке простых чисел, меньших 256, — 80%.

1.23.4. $N - 1$ методы доказательства простоты

Рассмотрим методы, позволяющие получить строгое доказательство простоты числа.

Для начала приведем хорошо известную теорему Люка, позволяющую доказать простоту числа n, если известно разложение числа $n - 1$ на простые множители.

Теорема 1.62 (**Люк**). Пусть $n > 1$ — нечетное натуральное число, $n-1 = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ — каноническое разложение числа $n-1$. Если найдется такое целое число a, $(a, n) = 1$, для которого выполнены условия:

$$a^{n-1} \equiv 1 \pmod{n}, \quad a^{\frac{n-1}{p_i}} \not\equiv 1 \pmod{n}, \quad i = 1, \dots, k,$$

то n — простое число.

Доказательство. Пусть существует такое целое a, для которого выполнены условия теоремы. Тогда из теорем 1.17 и 1.41 следует, что $P_n(a) = n - 1$, где $P_n(a)$ — показатель числа a по

модулю n. Так как $P_n(a) \leqslant \varphi(n) \leqslant n-1$, то $\varphi(n) = n-1$, а это возможно только в случае, когда n — простое число. $\square$

Основной сложностью в теореме Люка является знание полного разложения числа $n-1$ на простые множители. Такое на практике возникает далеко не часто.

Рассмотрим методы доказательства простоты числа n, когда известно лишь частичное разложение числа $n-1$ на множители. Пусть известно разложение:

$$n-1 = fr, \quad (f,r) = 1, \quad f = q_1^{\alpha_1} \dots q_s^{\alpha_s} -$$

каноническое разложение числа f, r — составное число с неизвестным разложением на множители. Заметим, что если r — простое, то мы получим полное разложение числа $n-1$, что позволяет воспользоваться теоремой Люка для проверки числа n на простоту.

Также будем считать, что каждый простой делитель q числа r удовлетворяет неравенству $q > B$ для некоторого натурального B. Например, в качестве B можно взять значение:

$$B = \max_{k=1}^{s} q_k.$$

Определим следующие условия.

1. Для каждого простого числа q_k, $k = 1, \dots, s$, входящего в разложение числа f, найдется некоторое целое число a_k, $(a_k, n) = 1$, такое, что:

$$a_k^{n-1} \equiv 1 \pmod{n}, \quad \left(a_k^{\frac{n-1}{q_k}} - 1, n\right) = 1. \tag{1.67}$$

2. Найдется некоторое целое число b, $(b,n) = 1$, такое, что:

$$b^{n-1} \equiv 1 \pmod{n}, \quad \left(b^{\frac{n-1}{r}} - 1, n\right) = 1. \tag{1.68}$$

Выполнимость одного из двух условий позволяет выносить суждения о простоте числа n.

Теорема 1.63 (**Поклингтон**). Пусть нечетное натуральное число n удовлетворяет условию (1.67). Тогда для любого простого делителя p числа n выполнено сравнение:

$$p \equiv 1 \pmod{f}.$$

Доказательство. Пусть выполнены условия теоремы и p — произвольный простой делитель числа n. Зафиксируем простой делитель q_k числа f. Пусть $P(a_k)$ — показатель числа a_k по модулю p.

Так как из сравнения $a_k^{n-1} \equiv 1 \pmod{n}$ следует сравнение $a_k^{n-1} \equiv 1 \pmod{p}$, то $P(a_k) | n-1$ (теорема 1.41). А из условия:

$$\left(a_k^{\frac{n-1}{q_k}} - 1, n\right) = 1$$

следует, что $a_k^{\frac{n-1}{q_k}} \not\equiv 1 \pmod{p}$. Поэтому $P(a_k)$ не является делителем числа $\dfrac{n-1}{q_k}$. Из условия $q_k^{\alpha_k} | n-1$ следует, что $q_k^{\alpha_k} | P(a_k)$.

С другой стороны, из малой теоремы Ферма следует, что $a_k^{p-1} \equiv 1 \pmod{p}$. Поэтому $P(a_k) | p-1$ и $q_k^{\alpha} | p-1$, что равносильно:

$$p \equiv 1 \pmod{q_k^{\alpha_k}}.$$

В силу произвольного выбора q_k последнее сравнение выполнено для любого $k = 1, \ldots, s$. Поэтому по китайской теореме об остатках 1.33 выполнено сравнение $p \equiv 1 \pmod{f}$, так как $f = q_1^{\alpha_1} \ldots q_s^{\alpha_s}$. □

Теорему Поклингтона можно применять для доказательства простоты числа.

Теорема 1.64 (**Лемер, следствие теоремы Поклингтона**). Если нечетное натуральное n удовлетворяет условию (1.67) и $f \geqslant \sqrt{n}$, то n — простое.

Доказательство. Из теоремы 1.63 следует, что для любого простого делителя p числа n найдется такое t, зависящее от p, что $p = 1 + ft$. Так как $t \geqslant 1$ и $f \geqslant \sqrt{n}$, то:

$$p = 1 + ft \geqslant 1 + \sqrt{n} > \sqrt{n}.$$

Учитывая предложение 1.9, получаем, что n — простое. $\square$

Следствие 1.17. Предположим, что у числа $n-1$ имеется простой делитель q, $q \geqslant \sqrt{n}$, а также существует такое $a \in \mathbb{Z}$, $(a, n) = 1$, для которого выполнены следующие условия:

$$a^{n-1} \equiv 1 \pmod{n}, \quad (a^{(n-1)/q} - 1,\ n) = 1.$$

Тогда n является простым.

Согласно Нилу Коблицу, значение $a = 2$ часто подходит для проверки с помощью следствия 1.17.

Пример 1.20. Рассмотрим число $n = 89$. Простое число $p = 11$ является делителем числа 88 и $11 \geqslant \sqrt{89}$. Число $a = 2$ удовлетворяет условиям:

$$2^{88} \equiv 1 \pmod{89}, \quad \left(2^8 - 1, 89\right) = 1.$$

Поэтому согласно следствию 1.17 число $n = 89$ является простым.

Пример 1.21. Покажем, как можно на основе следствия 1.17 генерировать простые числа. Предположим, что нужно найти простое число длины 10 десятичных знаков.

Выберем произвольное простое число $q_1 \geqslant 5$. Пусть $q_1 = 17$. Выберем четное число r с условием $2 \leqslant r \leqslant q_1 - 3$. Пусть $r = 14$. Рассмотрим число $n = q_1 r + 1 = 239$. Для того, чтобы число n было простым, достаточно найти такое целое a, $(a, n) = 1$, для которого $a^{n-1} \equiv 1 \pmod{n}$ и $(a^r - 1, n) = 1$. Этим условием удовлетворяет, например, $a = 2$.

Обозначим $q_2 = 239$. Выберем четное число r с условием $2 \leqslant r \leqslant q_2 - 3$. Пусть $r = 236$ и $n = q_2 r + 1$. Так как при $a = 2$, $(a, n) = 1$, $a^{n-1} \not\equiv 1 \pmod{n}$, то число n — составное. Поэтому выберем другое r, например, $r = 228$, $n = q_2 r + 1 = 54493$. При $a = 2$ условия следствия 1.17 выполнены, поэтому $n = 54493$ — простое.

Обозначим $q_3 = 54493$. Выберем четное число r с условием $2 \leqslant r \leqslant q_3 - 3$. Пусть $r = 21112$ и $n = q_3 r + 1 = 1150456217$. При

$a = 2$ условия следствия выполнены, поэтому $n = 1150456217$ — простое число требуемой длины.

При этом заметим, что:

$$n = q_k r + 1 \leqslant q_k(q_k - 3) + 1 = q_k^2 - 3q_k + 1 \leqslant q_k^2 - 3 \cdot 5 + 1 < q_k^2.$$

Теперь покажем, как можно использовать условие (1.68) для доказательства простоты.

Теорема 1.65. Пусть n — нечетное натуральное число, для которого выполнено условие (1.68). Тогда для любого простого делителя p числа n найдется такой простой делитель q числа r, зависящий от p, что $p \equiv 1 \pmod{q}$.

Доказательство. Пусть p — произвольный простой делитель числа n, $P(b)$ — показатель числа b по модулю p.

Так как из сравнения $b^{n-1} \equiv 1 \pmod{n}$ следует $b^{n-1} \equiv 1 \pmod{p}$, то $P(b) | n - 1$ (теорема 1.41). А из условия:

$$\left(b^{\frac{n-1}{r}} - 1, n\right) = 1$$

следует, что $b^{\frac{n-1}{r}} \not\equiv 1 \pmod{p}$. Поэтому $P(b)$ не является делителем числа $\dfrac{n-1}{r}$. Следовательно, $(P(b), r) > 1$ и для некоторого простого делителя q числа r выполнено $q | P(b)$.

С другой стороны, из малой теоремы Ферма следует, что $b^{p-1} \equiv 1 \pmod{p}$. Поэтому $P(b) | p - 1$ и $q | p - 1$, что равносильно $p \equiv 1 \pmod{q}$. □

Теорема 1.66. Пусть n — нечетное натуральное число, для которого выполнены условия (1.67) и (1.68). Тогда для любого простого делителя p числа n найдется такой простой делитель q числа r, зависящий от p, что $p \equiv 1 \pmod{fq}$.

Доказательство следует из китайской теоремы об остатках и теорем 1.63 и 1.65. □

Следствие 1.18. Пусть n — нечетное натуральное число, для которого выполнены условия (1.67), (1.68) и $fB \geqslant \sqrt{n}$. Тогда число n является простым.

1.24. Разложение целых чисел на множители

Пусть n — натуральное составное число. Требуется найти натуральные числа n_1, n_2, большие единицы и меньшие n, для которых $n = n_1 n_2$. Данная задача носит название *задачи факторизации* числа n. Если какое-то из чисел n_1 или n_2 составное, то к нему также применяют процедуру поиска делителей. В конечном итоге, для числа n находится его каноническое разложение: $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$.

Задача фаркторизации является более сложной, нежели проверка числа на простоту. Поэтому, прежде чем раскладывать число на множители, его проверяют на простоту.

1.24.1. Метод пробного деления

Самым простым и очевидным алгоритмом поиска делителей числа n является метод пробного деления. Он заключается в последовательном делении числа n на числа, не превосходящие значения $[\sqrt{n}]$. Данная оценка сверху верна в силу предложения 1.9: наименьший простой делитель числа n не превосходит $\sqrt{n}$. С теоретической точки зрения, число n достаточно делить только на простые числа. Для этого необходимо иметь таблицу простых чисел от 2 до $[\sqrt{n}]$, которую можно построить, например, с помощью решета Эратосфена. Так как одно пробное деление n на p требует не более $O(\log^2 n)$ операций, а количество пробных делений равно $\pi(n) = O\left(\dfrac{\sqrt{n}}{\ln n}\right)$ (теорема Чебышева), то трудоемкость всего алгоритма оценивается величиной $O(\sqrt{n}\log n)$. С практической же точки зрения, такая таблица может занимать очень много места в памяти компьютера или же вовсе не поместиться в ней. При делении числа n на все числа от 2 до $\sqrt{n}$ трудоемкость алгоритма возрастет в $\log n$ раз и составит $O(\sqrt{n}\log^2 n)$.

На практике строится таблица простых чисел в небольшом диапазоне (например, до 2^{16}) и проверка производится только

для этих делителей. Поиск больших простых делителей выполняется другими алгоритмами.

1.24.2. ρ-метод Полларда

Одним из наиболее популярных алгоритмов факторизации является ρ-метод Полларда, предложенный в 1975 г.

Идея данного метода заключается в следующем. Пусть X — конечное множество из n элементов, $f : X \to X$, $x_0 \in X$ и построена рекуррентная последовательность $x_{i+1} = f(x_i)$, $i \geqslant 0$. Нетрудно видеть, что последовательность $\{x_i\}$ является периодической и ее период не превосходит n.

В данном случае в качестве множества X берется кольцо вычетов $\mathbb{Z}_n$, f — некоторый многочлен с целыми коэффициентами, степени большей единицы. В своей работе Полларда предложил многочлен $f(x) = x^2 + 1$.

Пусть N и t — такие натуральные числа, что для любого $k \geqslant N$ выполнено сравнение:

$$x_{k+t} \equiv x_k \pmod{n}.$$

Величина t называется периодом последовательности $\{x_i\}$, N — длиной подхода к периоду. При этом, скорее всего, длина периода t равна n.

Так как n — составное число, то найдется его простой делитель $p \leqslant \sqrt{n}$. Из последнего сравнения следует, что:

$$x_{k+t} \equiv x_k \pmod{p}.$$

При этом длина периода последовательности $\{x_i \pmod{p}\}$ не превосходит числа p. Значит, с большой вероятностью найдутся такие x_i, x_j, что $x_i \not\equiv x_j \pmod{n}$, но $x_i \equiv x_j \pmod{p}$. Последние условия означают, что $1 < (x_i - x_j, n) < n$. Таким образом, метод Полларда сводится к поиску таких x_i, x_j, для которых $x_i \equiv x_j \pmod{p}$. Так как число p неизвестно, то все вычисления в алгоритме проводятся по модулю n и на каждом шаге вычисляется $d = (x_i - x_j, n)$. Нетривиальный наибольший общий делитель $1 < d < n$ как раз и будет искомым делителем числа

n. Случай $d = n$ имеет место с пренебрежимо малой вероятностью.

Для поиска необходимых элементов x_i, x_j Полларда предложил использовать метод Флойда поиска циклов в последовательностях: вычислить две последовательности $\{x_i\}$ и $\{x_{2i}\}$ и находить простой делитель p проверкой условия:

$$(z, n) > 1, \quad z \equiv x_{2i} - x_i \pmod{n}.$$

Алгоритм 1.10 (**ρ-метод Полларда**).

Вход: целое составное n.

Выход: нетривиальный делитель d числа n $(1 < d < n)$.

1. Зафиксировать некоторый многочлен второй степени $f(x) \in \mathbb{Z}[x]$, например, $f(x) = x^2 + 1$.

2. Выбрать случайное целое число x_0, $0 < x_0 < n$, и определить $a := x_0$, $b := x_0$.

3. Вычислить:

$$a := f(a) \pmod{n}, \quad b := f(b) \pmod{n}, \quad b := f(b) \pmod{n}.$$

4. Вычислить $d = (a - b, n)$.

5. Если $d > 1$, то завершить работу и возвратить d. В противном случае вернуться в шаг 3.

Трудоемкость данного метода составляет $O(\sqrt{p}) \sim O(\sqrt[4]{n})$.

1.24.3. $(p-1)$-метод Полларда

Данный метод особенно эффективен при разложении чисел частного вида.

Пусть $B = \{p_1, p_2, \ldots, p_k\}$ — множество различных простых чисел (база разложения). Целое число называется *B-гладким*, если все его простые делители являются элементами множества B.

Пусть n — нечетное составное натуральное число, которое требуется разложить на множители и p — некоторый простой делитель числа n, $p-1 = p_1^{\alpha_1} \ldots p_k^{\alpha_k}$ — каноническое разложение числа $p-1$. Найдем максимальные показатели $r_1, r_2, \ldots, r_k$, для

которых выполнено неравенство $p_i^{r_i} \leqslant n$, $i = 1, \ldots, k$. Логарифмируя обе части данного неравенства, получаем $r_i \ln p_i \leqslant \ln n$, откуда $r_i \leqslant \left[\dfrac{\ln n}{\ln p_i}\right]$. Обозначим:

$$M = p_1^{\left[\frac{\ln n}{\ln p_i}\right]} \ldots p_k^{\left[\frac{\ln n}{\ln p_k}\right]}.$$

Тогда для некоторого целого z выполнено $M = (p-1)z$.

Пусть a — некоторое целое число, взаимно простое с p. Согласно малой теореме Ферма, выполнено $a^{p-1} \equiv 1 \pmod{p}$. Тогда $a^M \equiv 1 \pmod{p}$, так как:

$$a^M \equiv (a^{p-1})^z \equiv 1 \pmod{p}.$$

Данное сравнение позволяет заключить, что:

$$p \,|\, (n,\ a^M - 1 \pmod{n}).$$

Алгоритм 1.11 (**$(p-1)$-метод Полларда**).

Вход: целое составное n.

Выход: нетривиальный делитель d числа n.

1. Выбрать базу разложения $B = \{p_1, p_2, \ldots, p_k\}$.

2. Выбрать случайное целое число a, $1 < a < n-1$.

3. Вычислить $d = (a, n)$. Если $d > 1$, то завершить алгоритм и возвратить d.

4. Цикл $i = 1, \ldots, k$.

   4.1. Вычислить $r := \left[\dfrac{\ln n}{\ln p_i}\right]$.

   4.2. Положить $a := a^r \pmod{n}$.

5. Вычислить $d := (a-1, n)$.

6. Если $1 < d < n$, то возвратить d. Если же $d = 1$ или $d = n$, то завершить алгоритм с уведомлением о том, что делитель не найден.

Заметим, что если среди простых делителей числа n имеются такие делители, что число $p-1$ является B-гладким, и имеются такие делители, что число $p-1$ не является B-гладким, то алгоритм 1.11 найдет нетривиальный делитель.

Если для всех простых делителей p числа n число $p-1$ не является B-гладким, то для любого a в алгоритме 1.11 будет получиться результат $(a-1,n)=1$.

Если для всех простых делителей p числа n число $p-1$ является B-гладким, то в алгоритме 1.11 может получиться результат $(a-1,n)=n$.

1.25. Методы дискретного логарифмирования

Пусть $(G,\cdot)$ — конечная циклическая группа порядка m, g — образующий элемент G и $y \in G$. Дискретным логарифмом (показателем) элемента y группы G по основанию g называется число $x \in \{0,1,\ldots,m-1\}$, являющееся решением уравнения:

$$g^x = y.$$

Дискретный логарифм элемента y по основанию g обозначается $\log_g y$. Если групповая операция задана в аддитивной форме, то данное уравнение записывается в виде $xg = y$.

Для криптографических протоколов наиболее важными являются следующие циклические группы:

1. Мультипликативная группа $\mathbb{Z}_p^*$ кольца вычетов по простому модулю p.

2. Мультипликативная группа $GF(q)^*$ конечного поля из $q = p^n$ элементов.

3. Циклическая подгруппа точек эллиптической кривой $E(GF(q))$ над полем $GF(q)$.

В данном параграфе, для определенности, рассматривается первый случай. Запишем его в следующем виде. Пусть p — некоторое простое число, g — некоторый первообразный корень по модулю p. Хорошо известно, что для любого натурального x можно вычислить значение $y = g^x \pmod p$, затратив на это не более $2\cdot\log_2 p$ операций (умножения). Зададимся обратной задачей: пусть известны p, g и y, требуется отыскать (единственное) значение x, $0 < x < p$, при котором $y = g^x \pmod p$.

Пусть t_y — число операций (умножения), необходимых для вычисления $y = g^x \pmod p$. Тогда $t_y \leqslant 2\cdot\log_2 p$. Обозначим через

n величину $\log_2 p$ (количество бит в числе p). Тогда получим такое неравенство: $t_y \leqslant 2n$, т.е. сложность вычисления y от x ограничена сверху линейной функцией от n.

1.25.1. Полный перебор

Будем перебирать все значения x, $0 < x < p$, пока $g^x \pmod p$ не будет равно y. Пусть x является случайной величиной, которая имеет равномерный закон распределения на множестве $\{1, 2, \ldots, p-1\}$, т.е. для любого $i = 1, \ldots, p-1$ выполнено такое равенство: $P(x = i) = \dfrac{1}{p-1}$. Тогда математическое ожидание данной случайной величины равно следующему значению:

$$M(x) = \sum_{i=1}^{p-1} i \cdot P(x = i) = \frac{1}{p-1} \sum_{i=1}^{p-1} i = \frac{p}{2}.$$

Поэтому в среднем потребуется в уравнение $y = g^x \pmod p$ подставить $p/2$ значений x, чтобы найти решение. Заметим, что при подстановке значения $x = 1$ требуется 0 операций умножения, при подстановке $x = 2$ требуется 1 операция умножения: $g^2 = g \cdot g$ и т.д. Каждый раз запоминая результат предыдущего шага, получаем $g^i = g^{i-1} \cdot g$, поэтому для проверки $x = i$ требуется $i - 1$ операций умножения.

Пусть φ — случайная величина, отвечающая за сложность операции при вычислении $g^i \pmod p$, $i = 1, \ldots, p-1$. Данная случайная величина определена на множестве $\{0, 1, \ldots, p-2\}$, причем:

$$P(\varphi = i) = P(x = i + 1) = \frac{1}{p-1}, \quad i = 0, \ldots, p-2.$$

Поэтому:

$$M(\varphi) = \sum_{i=0}^{p-2} i \cdot P(\varphi = i) = \frac{p-2}{2} \approx \frac{p}{2}.$$

Таким образом, сложность полного перебора в среднем приближенно равна $p/2$. Если выразить это значение через n (длина двоичного представления числа p), то сложность в среднем

приближенно равна 2^{n-1}, т.е. сложность данного метода растет экспоненциально относительно n.

1.25.2. Метод Гельфонда-Шенкса

Попробуем усовершенствовать метод полного перебора. Зафиксируем некоторые натуральные числа a и b с тем условием, что $ab > p$. Из теоремы 1.1 следует, что найдется, и притом единственная, пара чисел q и r, что $x = qa + r$, $0 \leqslant r < a$, причем в данном случае число q будет неотрицательным. Рассмотрим две конечные последовательности:

$$y,\quad gy,\quad g^2y,\ldots,\quad g^{a-1}y \pmod p, \tag{1.69}$$

$$g^a,\quad g^{2a},\quad g^{3a},\ldots,\quad g^{ba} \pmod p. \tag{1.70}$$

Предложение 1.57. В последовательности (1.69) найдется такой элемент, который равен хотя бы одному элементу последовательности (1.70).

Доказательство. Представим каждый элемент последовательности (1.69) в виде:

$$g^iy = g^ig^{qa+r} = g^{qa}g^{i+r},\quad i = 0,\ldots,a-1.$$

Относительно чисел q и r возможны следующие случаи.

1. $q = 0$. Тогда $x < a$. Так как $x \geqslant 1$, то в данном случае $1 \leqslant r = x < a$. Элементы последовательности (1.69) имеют вид:

$$g^iy = g^{i+r},\quad i = 0,\ldots,a-1.$$

Так как $0 \leqslant i \leqslant a-1$ и $1 \leqslant r < a$, то $1 \leqslant i+r < 2a-1$. Поэтому найдется такое значение i, что $i+r = a$. При этом $g^iy = g^a$, что соответствует первому элементу последовательности (1.70).

2. $r = 0$. Тогда $q \geqslant 1$ и:

$$g^iy = g^{qa}g^i,\quad i = 0,\ldots,a-1.$$

В этом случае утверждение предложения выполнено при $i = 0$.

3. $q \geqslant 1$, $r \geqslant 1$. Тогда $a > 1$ и $1 \leqslant i + r \leqslant 2a - 2$. Поэтому найдется такое i, что $i + r = a$ и будет выполнено равенство:

$$g^iy = g^{qa}g^{i+r} = g^{(q+1)a}.$$ □

Таким образом, найдутся такие i и j, где $0 \leqslant i \leqslant a-1$, $1 \leqslant j \leqslant b$, что $g^i y = g^{ja}$:

$$g^{ja} = g^i y = q^i g^x = g^{i+x} \pmod{p}.$$

Из данного равенства видно, что $x \equiv ja - i \pmod{p-1}$.

Известно, что сложность данного алгоритма равна $O(n \cdot 2^{n/2})$. Недостатком данного метода является хранение одной из последовательностей 1.69 и 1.70 в памяти компьютера.

Пример 1.22. Пусть $p = 23$ и $g = 5$. Заметим, что в данном случае выполнены все условия теоремы 1.45: $22 = 2 \cdot 11$:

$$5^2 \equiv 2 \pmod{23}, \quad 5^{11} \equiv 22 \pmod{23}.$$

Поэтому число $g = 5$ является первообразным корнем по модулю $p = 23$:

$$\{2^x \pmod{23} |\ x = 1, \ldots, 22\} = \{1, 2, \ldots, 22\}.$$

Пусть требуется решить сравнение $5^x \equiv 17 \pmod{23}$. Зафиксируем числа $a = 6$, $b = 4$. Получаем следующие значения элементов последовательностей (1.69) и (1.70):

$$17, \quad 16, \quad 11, \quad 9, \quad 22, \quad 18 \pmod{23},$$

$$8, \quad 18, \quad 6, \quad 2 \pmod{23}.$$

При $i = 5$, $j = 2$ выполнено равенство $5^5 \cdot 17 = 5^{2 \cdot 6} \pmod{23}$. Поэтому $x = 2 \cdot 6 - 5 = 7$ является решением исходного сравнения.

1.25.3. ρ-метод Полларда

Рассмотрим сравнение:

$$a^x \equiv b \pmod{p}, \quad P(a) = m, \tag{1.71}$$

где $P(a)$ — показатель числа a по модулю p.

Рассмотрим конечное множество возможных степеней элемента a :

$$A = \{1, a, a^2, \ldots, a^{m-1} \pmod{p}\}.$$

Для некоторого фиксированного числа $s \geqslant 3$ зафиксируем разбиение множества A на s частей:

$$A = I_0 \cup I_1 \cup \ldots \cup I_{s-1},$$

где $I_i \cap I_j = \varnothing$ при $i \neq j$. Например, можно определить $z \in A$ принадлежит множеству I_i, если $z \equiv i \pmod{s}$.

Для всех значений $i = 0, 1, \ldots, s-1$ зафиксируем значения α_i, $\beta_i \in \mathbb{Z}_m$, однозначно связанные с множеством I_i. Определим отображение множества A в себя $f : A \to A$ следующим образом:

$$f(z) = \begin{cases} za^{\alpha_0}b^{\beta_0} \pmod{p}, & z \in I_0, \\ za^{\alpha_1}b^{\beta_1} \pmod{p}, & z \in I_1, \\ \ldots & \ldots \\ za^{\alpha_{s-1}}b^{\beta_{s-1}} \pmod{p}, & z \in I_{s-1}, \end{cases} \tag{1.72}$$

где числа a и b определены в (1.71).

Заметим, что если $z \in A$, то $z \equiv a^{\gamma} \pmod{p}$ для некоторого γ. Поэтому:

$$f(z) = a^{\gamma+\alpha_i+x\beta_i} \pmod{p}.$$

Это объясняет корректность определения отображения f.

Зафиксируем случайным образом число k_0, $0 < k_0 < m$, и определим $z_0 \in A$ сравнением $z_0 \equiv a^{k_0} \pmod{p}$. Рассмотрим рекуррентную последовательность $\{z_n\}_{n\geqslant 0}$, определенную равенством:

$$z_{n+1} = f(z_n), \quad n \geqslant 0.$$

Учитывая определение функции f, получаем:

$$z_{n+1} \equiv z_n a^{\alpha_i} b^{\beta_i} \equiv a^{A_{n+1}+xB_{n+1}} \pmod{p}, \quad z_n \in I_i,$$

где A_{n+1}, B_{n+1} определяются равенствами:

$$A_0 = k_0, \quad B_0 = 0, \quad A_{n+1} = A_n + \alpha_i, \quad B_{n+1} = B_n + \beta_i.$$

Так как множество A конечно, то последовательность $\{z_n\}_{n\geqslant 0}$ зациклится. При этом найдутся такие индексы n и r, для которых $z_n \equiv z_r \pmod{p}$:

$$a^{A_n+xB_n} \equiv a^{A_r+xB_r} \pmod{p}.$$

Так как $P(a) = m$, то:

$$A_n + xB_n \equiv A_r + xB_r \pmod{m}.$$

Следовательно:

$$x \equiv \frac{A_n - A_r}{B_r - B_n} \pmod{m}.$$

Для обнаружения сравнения $z_n \equiv z_r \pmod{p}$ Поллард предложил использовать метод Флойда определения циклов в последовательностях, т.е. в данном случае проверять, выполнено ли сравнение $z_n \equiv z_{2n} \pmod{p}$ для некоторого n.

Алгоритм 1.12 (**ρ-метод Полларда**).

Вход: простое число p, целые числа a, b, большие нуля и меньшие p, удовлетворяющие сравнению $a^x \equiv b \pmod{p}$, где $P(a) = m$, а также параметр $s \geqslant 3$ и отображение f, задаваемая наборами чисел α_i, $\beta_i \in \mathbb{Z}_m$, $i = 0, 1, \ldots, s-1$ (1.72).

Выход: дискретный логарифм $x \equiv \log_a b \pmod{m}$.

1. Выбрать случайное k_0, для которого $0 < k_0 < m$, и определить $z_{2n} := z_n := a^{k_0} \pmod{p}$.

2. Определить начальные значения:

$$A_{2n} := A_n := k_0, \quad B_{2n} := B_n := 0.$$

3. Вычислить $i \equiv z_n \pmod{s}$, $z_n = f(z_n)$ и определить:

$$A_n := A_n + \alpha_i \pmod{m}, \quad B_n := B_n + \beta_i \pmod{m}.$$

4. Вычислить $i \equiv z_{2n} \pmod{s}$, $z_{2n} = f(z_{2n})$:

$$A_{2n} := A_{2n} + \alpha_i \pmod{m}, \quad B_{2n} := B_{2n} + \beta_i \pmod{m},$$

$$j \equiv z_{2n} \pmod{s}, \quad z_{2n} = f(z_{2n}),$$

$$A_{2n} := A_{2n} + \alpha_j \pmod{m}, \quad B_{2n} := B_{2n} + \beta_j \pmod{m}.$$

5. Если $z_n \neq z_{2n} \pmod{m}$, то вернуться в шаг 3.

6. Если $A_n = A_{2n}$ или $B_n = B_{2n}$, то вернуться в шаг 3.

7. Определить $x : x \equiv \dfrac{A_n - A_{2n}}{B_{2n} - B_n} \pmod{m}$.

Сложность данного алгоритма составляет $O(\sqrt{m})$. ρ-метод Полларда имеет ту же сложность, что и метод Гельфанда-Шенкса, но использует лишь ограниченное количество памяти компьютера.

1.25.4. Метод исчисления порядка

Рассмотрим еще один метод нахождения неизвестного значения x, $0 < x < p$, по известному y, $0 < y < p$.

Для начала заметим, что множество $\mathbb{Z}_p^* = \mathbb{Z}_p \backslash \{0\}$ является мультипликативной группой порядка $p-1$, поэтому будет выполнено такое равенство:

$$\{y \cdot g^x \pmod p \mid x = 1, \ldots, p-1\} = \{1, \ldots, p-1\}.$$

Также нам понадобится следующий факт: если случайным образом из множества целых чисел выбрать некоторое число, то с вероятностью $1/z$ данное число будет делиться на z, где z — некоторое натуральное число. Поэтому чем больше число z, тем меньше такая вероятность. Исходя из двух данных фактов строится алгоритм исчисления порядка.

1. Первым шагом фиксируем «достаточно» большое множество $M_n = \{p_1, \ldots, p_n\}$ первых n простых чисел.

2. Последовательно перебираем элементы вида $g^k \pmod p$, $k = 1, 2, \ldots$, и учитываем те из них, у которых при разложении на простые множители не встречаются простые числа, большие числа p_n :

$$g^{k_i} = p_1^{a_{i1}} \ldots p_n^{a_{in}} \pmod p, \quad a_{i1}, \ldots, a_{in} \in \mathbb{Z}. \tag{1.73}$$

   Как только количество таких чисел $k_1, k_2, \ldots$ будет $n+\varepsilon$, где ε — небольшое натуральное число (при больших значениях p число ε можно зафиксировать в диапазоне от 0 до 10), то процесс перебора значений k завершается. Логарифмируя равенства (1.73), получаем такие сравнения:

$$k_i \equiv a_{i1} \cdot \log_g p_1 + \ldots + a_{in} \cdot \log_g p_n \pmod{p-1}.$$

3. Обозначив

$$x_i = \log_g p_i, \quad i = 1, \ldots, n,$$

   получаем такую систему линейных уравнений относительно

неизвестных $x_1,\ldots,x_n$:

$$\begin{cases} a_{11}\cdot x_1+\ldots+a_{1n}\cdot x_n & \equiv \quad k_1 \pmod{p-1}, \\ & \ldots \\ a_{n+\varepsilon,1}\cdot x_1+\ldots+a_{n+\varepsilon,n}\cdot x_n & \equiv \quad k_{n+\varepsilon} \pmod{p-1}. \end{cases} \tag{1.74}$$

Число ε здесь необходимо из тех соображений, что если рассматривать ровно n уравнений, то полученная система может оказаться неопределенной, т.е. иметь более одного решения. Поэтому количество уравнений в системе выбирается с некоторым «запасом». Если все же система оказалась неопределенной, то ее необходимо дополнить уравнениями, которые составлялись во втором пункте данного алгоритма. Алгоритм решения системы сравнений описывается в следующем параграфе.

4. Пусть после решения системы (1.74) значения $x_i=\log_g p_i$, $i=1,\ldots,n$, найдены. Выберем случайным образом некоторое целое r, $1\leqslant r\leqslant p-1$. Исходя из приведенных двух фактов в начале параграфа следует, что «достаточно» мала вероятность того, что в разложении числа $y\cdot g^r \pmod p$ на простые множители присутствует простое число, большее числа p_n (вероятность данного события меньше, чем $1/p_n$). Если же такое простое число присутствует, то выберем другое число r.

Итак, пусть $y\cdot g^r = p_1^{b_1}\ldots p_n^{b_n} \pmod p$. Запишем данное равенство в таком виде:

$$g^x\cdot g^r = g^{x+r} = p_1^{b_1}\ldots p_n^{b_n} \pmod p.$$

Логарифмируя последнее равенство, получаем:

$$x\equiv b_1\cdot\log_g p_1+\ldots+b_n\cdot\log_g p_n - r \pmod{p-1}.$$

Алгоритм исчисления порядка является одним из наиболее эффективных и быстродействующих алгоритмов решения задачи дискретного логарифмирования при не очень больших p. Тем не менее, его вычислительная сложность является достаточно высокой.

Известно, что сложность данного алгоритма сверху оценивается следующим числом:

$$c_1 2^{(c_2+o(1))\sqrt{n\cdot\log n}},$$

где c_1, c_2 — некоторые положительные константы.

Пример 1.23. Пусть $p = 47$ и $g = 5$. Решим методом исчисления порядка уравнение $5^x \equiv 17 \pmod{47}$.

1. Пусть $M_4 = \{2, 3, 5, 7\}$.

2. Пусть $\varepsilon = 1$.

$$\begin{aligned}
5^1 \pmod{47} &= 5 &&= 5, && \surd\\
5^2 \pmod{47} &= 25 &&= 5\cdot 5, && \surd\\
5^3 \pmod{47} &= 31 &&= 31, &&\\
5^4 \pmod{47} &= 14 &&= 2\cdot 7, && \surd\\
5^5 \pmod{47} &= 23 &&= 23, &&\\
5^6 \pmod{47} &= 21 &&= 3\cdot 7, && \surd\\
5^7 \pmod{47} &= 11 &&= 11, &&\\
5^8 \pmod{47} &= 8 &&= 2^3. && \surd
\end{aligned}$$

3. Пусть $x_1 = \log_5 2$, $x_2 = \log_5 3$, $x_3 = \log_5 5$, $x_4 = \log_5 7$. Получаем такую систему линейных уравнений:

$$\begin{cases}
x_3 \equiv & 1 \pmod{46},\\
2x_3 \equiv & 2 \pmod{46},\\
x_1 + x_4 \equiv & 4 \pmod{46},\\
x_2 + x_4 \equiv & 6 \pmod{46},\\
3x_1 \equiv & 8 \pmod{46}.
\end{cases}$$

Решив данную систему, получаем:

$$x_1 \equiv 18,\ x_2 \equiv 20,\ x_3 \equiv 1,\ x_4 \equiv 32 \pmod{46}.$$

4. Рассмотрим $r = 1$. Тогда $17\cdot 5^1 \pmod{47} = 38 = 2\cdot 19$. Данное число нам не подходит. Пусть $r = 2$. Тогда $17 \cdot 5^2 \pmod{47} = 2$. Данное число нам подходит, и:

$$x \equiv \log_5 2 - 2 = 16 \pmod{46}.$$

1.25.5. Решение систем линейных сравнений

Рассмотрим вопрос о решении систем сравнений, которые возникали в предыдущем параграфе. Будем следовать идеям работы [29]. Пусть имеется система сравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \equiv b_1 \pmod{p-1}, \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \equiv b_m \pmod{p-1}, \end{cases} \tag{1.75}$$

относительно неизвестных $x_1, \ldots, x_n$, $m \geqslant n$. При этом неизвестные представляют собой искомые индексы $\log_g p_1, \ldots, \log_g p_n$.

Предложение 1.58. Пусть $b_1, b_2 \in \mathbb{Z}$, одновременно не сравнимые с нулем по модулю $p-1$. Тогда найдутся такие целые числа $\alpha, \beta, \gamma, \delta$ такие, что $\alpha\delta - \beta\gamma = 1$ и выполнено равенство:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}, \quad z = (b_1, b_2).$$

Доказательство. В качестве α и β возьмем пару целых чисел, являющуюся решением диофантова уравнения $\alpha b_1 + \beta b_2 = z$ относительны неизвестных α, β, где $z = (b_1, b_2)$ (следствие 1.1). В качестве γ и δ зафиксируем такие значения:

$$\gamma = -\frac{b_2}{z}, \quad \delta = \frac{b_1}{z}.$$

Тогда:

$$\alpha\delta - \beta\gamma = \frac{1}{z}(\alpha b_1 + \beta b_2) = 1,$$

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \alpha b_1 + \beta b_2 \\ \gamma b_1 + \delta b_2 \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}. \qquad \square$$

Рассмотрим расширенную матрицу системы сравнений (1.75):

$$(A|B) = \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ & \ldots & & \ldots \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right).$$

Для фиксированных индексов i, j для матрицы $(A|B)$ определим преобразование $F_{ij}(\alpha, \beta, \gamma, \delta)$ следующим образом:

$$\left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ & \dots & & \dots \\ a_{i1} & \dots & a_{in} & b_i \\ & \dots & & \dots \\ a_{j1} & \dots & a_{jn} & b_j \\ & \dots & & \dots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right) \to \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ & \dots & & \dots \\ \alpha a_{i1} + \beta a_{j1} & \dots & \alpha a_{in} + \beta a_{jn} & \alpha b_i + \beta b_j \\ & \dots & & \dots \\ \gamma a_{i1} + \delta a_{j1} & \dots & \gamma a_{in} + \delta a_{jn} & \gamma b_i + \delta b_j \\ & \dots & & \dots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right)$$

по модулю числа $p-1$. Так как $\det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = 1$, то преобразование F_{ij} обратимо в кольце вычетов $\mathbb{Z}_{p-1}$. Поэтому данное преобразование не изменяет множество решений системы (1.75).

На основе предложения 1.58 приведем алгоритм приведения расширенной матрицы $(A|B)$ к ступенчатом виду с использованием преобразований вида F_{ij}.

Алгоритм 1.13.

Вход: Расширенная матрица $(A|B)$ системы (1.75).

Выход: Матрица $(A|B)$, приведенная к ступенчатому виду.

Цикл: i от 1 до $m-1$

Цикл: j от $i+1$ до m

Если $a_{ii} \not\equiv 0 \pmod{p-1}$ или $a_{ji} \not\equiv 0 \pmod{p-1}$, то для элементов a_{ii} и a_{ji} вычислить значения $\alpha, \beta, \gamma, \delta$ (предложение 1.58) и применить к матрице $(A|B)$ преобразование $F_{ij}(\alpha, \beta, \gamma, \delta)$. Полученную матрицу вновь обозначить через $(A|B)$.

После выполнения алгоритма 1.13 матрица $(A|B)$ примет вид:

$$(A|B) = \left(\begin{array}{ccccc|c} a_{11} & a_{12} & \dots & a_{1,n-1} & a_{1n} & b_1 \\ 0 & a_{22} & \dots & a_{2,n-1} & a_{2n} & b_2 \\ & \dots & & & \dots & \\ 0 & 0 & \dots & 0 & a_{rn} & b_r \\ 0 & 0 & \dots & 0 & 0 & b_{r+1} \\ & \dots & & & \dots & \\ 0 & 0 & \dots & 0 & 0 & b_m \end{array}\right),$$

где r — ранг матрицы A системы (1.75). Для нахождения дискретных логарифмов $\log_g p_1, \dots, \log_g p_n$ необходимо, чтобы $r = n$

и $b_{r+1} = \ldots = b_m = 0$. При этом заметим, что условие $b_{r+1} = \ldots = b_m = 0$ при решении задачи дискретного логарифмирования выполнено всегда, так как решение задачи существует в случае, когда g — первообразный корень. Предположим, что данные условия выполнены. Тогда для нахождения решения системы (1.75) последовательно находим x_n, $x_{n-1}, \ldots$, x_1 из сравнений

$$\begin{array}{ll} a_{nn}x_n \equiv b_n & (\mathrm{mod}\ p-1), \\ a_{n-1,n-1}x_{n-1} \equiv b_{n-1} - a_{n-1,n}x_n & (\mathrm{mod}\ p-1), \\ \ldots & \ldots \\ a_{11}x_1 \equiv b_1 - a_{12}x_2 - \ldots - a_{1n}x_n & (\mathrm{mod}\ p-1). \end{array}$$

Пример 1.24. Найдем решение системы сравнений:

$$\begin{cases} 2x + 5y + 4z \equiv 11 \pmod{30}, \\ 3x + y + 2z \equiv 1 \pmod{30}, \\ 4x + 2y + z \equiv 26 \pmod{30}. \end{cases}$$

Расширенная матрица, соответствующая данной системе, имеет вид:

$$(A|B) = \left(\begin{array}{ccc|c} 2 & 5 & 4 & 11 \\ 3 & 1 & 2 & 1 \\ 4 & 2 & 1 & 26 \end{array}\right).$$

Все преобразования над этой матрицей будем осуществлять по модулю 30. Сначала применим преобразование F_{12} с параметрами $\alpha = -1, \beta = 1, \gamma = -3, \delta = 2$. Получим матрицу:

$$(A|B) = \left(\begin{array}{ccc|c} 1 & 26 & 28 & 20 \\ 0 & 17 & 22 & 29 \\ 4 & 2 & 1 & 26 \end{array}\right).$$

Теперь к данной матрице применим преобразование F_{13} с параметрами $\alpha = -3, \beta = 1, \gamma = -4, \delta = 1$. Получим:

$$(A|B) = \left(\begin{array}{ccc|c} 1 & 14 & 7 & 26 \\ 0 & 17 & 22 & 29 \\ 0 & 28 & 19 & 6 \end{array}\right).$$

Наконец, применим к матрице $(A|B)$ преобразование F_{23} с параметрами $\alpha = 5, \beta = -3, \gamma = -28, \delta = 17$. Получим:

$$(A|B) = \left(\begin{array}{ccc|c} 1 & 14 & 7 & 26 \\ 0 & 1 & 23 & 7 \\ 0 & 0 & 7 & 10 \end{array}\right).$$

Данной матрице соответствует следующая система сравнений:

$$\begin{cases} x + 14y + 7z \equiv 26 \pmod{30}, \\ y + 23z \equiv 7 \pmod{30}, \\ 7z \equiv 10 \pmod{30}. \end{cases}$$

Из сравнения $7z \equiv 10 \pmod{30}$ находим $z \equiv 10 \pmod{30}$. Далее:

$$y \equiv 7 - 23 \cdot 10 \equiv 17 \pmod{30},$$

$$x \equiv 26 - 14 \cdot 17 - 7 \cdot 10 \equiv 18 \pmod{30}.$$

Глава 2. Алгебраические основы криптографии

2.1. Бинарные отношения

Пусть M — некоторое множество и $\mathcal{R}$ — некоторое подмножество декартова произведения $M \times M$. Говорят, что $\mathcal{R}$ определяет *бинарное отношение* на множестве M, при этом если $(x, y) \in \mathcal{R}$, то говорят, что элементы x и y находятся в отношении $\mathcal{R}$ (или связаны отношением $\mathcal{R}$), и обозначают $x\mathcal{R}y$.

Бинарное отношение $\mathcal{R}$ называют:

— *рефлексивным*, если $x\mathcal{R}x$ для любого $x \in M$;

— *симметричным*, если из $x\mathcal{R}y$ следует, что $y\mathcal{R}x$;

— *транзитивным*, если из $x\mathcal{R}y$ и $y\mathcal{R}z$ следует, что $x\mathcal{R}z$;

— *антисимметричным*, если из $x\mathcal{R}y$ и $y\mathcal{R}x$ следует $x = y$.

2.1.1. Отношение эквивалентности

Важным классом бинарных отношений являются бинарные отношения, обладающие свойствами рефлексивности, симметричности и транзитивности. Такие бинарные отношения носят названия *отношений эквивалентности*. Если на множестве M задано отношение эквивалентности, то это обозначают $(M, \sim)$, при этом если пара элементов $x, y \in M$ связана отношением эквивалентности, то говорят, что x и y эквивалентны, и обозначают $x \sim y$ (в некоторых случаях используются и другие обозначения, например $x = y$, $x \equiv y$).

Пример 2.1. 1. Пусть M — произвольное множество. Определим на множестве M отношение эквивалентности, определив $x \sim y$ (для элементов $x, y \in M$) тогда и только тогда, когда $x = y$.

2. Пусть $f : M \to N$ — некоторое отображение из M в N. Определим на множестве M отношение эквивалентности, положив $x \sim y$ (для элементов $x, y \in M$) тогда и только тогда, когда $f(x) = f(y)$.

3. Пусть $M_n(K)$ — множество всех квадратных матриц порядка n над полем K. Для матриц $A, B \in M_n(K)$ определим $A \sim B$ тогда и только тогда, когда найдется такая невырожденная матрица $C \in M_n(K)$, что будет верным равенство $B = C^{-1}AC$, где C^{-1} — обратная матрица к C.

4. Пусть m — некоторое фиксированное положительное целое число. На множества целых чисел $\mathbb{Z}$ определим отношение эквивалентности, положив $x \equiv y$ тогда и только тогда, когда разность $y - x$ делится нацело на число m. В этом случае числа x и y называются *сравнимыми по модулю* m и это обозначается $x \equiv y \pmod{m}$.

Пусть на множестве M задано отношение эквивалентности $\sim$ и x — некоторый элемент из M. Обозначим через $\overline{x}$ — подмножество в M, состоящее из всех таких элементов множества M, которые эквивалентны элементу x : $\overline{x} = \{y \in M \mid x \sim y\}$. При этом множество $\overline{x}$ называется *классом эквивалентности, порожденным элементом* x.

Предложение 2.1. Для любого элемента $y \in \overline{x}$ верно равенство $\overline{x} = \overline{y}$.

Доказательство. Пусть $y \in \overline{x}$. В силу определения $\overline{x}$ выполнено отношение $x \sim y$, из чего следует, что $y \sim x$. Рассмотрим любой элемент $z \in \overline{x}$. Тогда из $y \sim x$ и $x \sim z$ следует $y \sim z$. Поэтому $z \in \overline{y}$ и $\overline{x} \subseteq \overline{y}$.

Обратно, пусть $z \in \overline{y}$. Тогда из $x \sim y$ и $y \sim z$ следует $x \sim z$, поэтому $z \in \overline{x}$ и $\overline{y} \subseteq \overline{x}$. Таким образом, $\overline{x} = \overline{y}$. □

Предложение 2.2. Любые два класса эквивалентности (определенные на одном множестве) либо не пересекаются, либо равны.

Доказательство. Пусть $\overline{x}$ и $\overline{y}$ — два некоторых класса эк-

вивалентности. Предположим, что $\overline{x} \cap \overline{y} \neq \varnothing$. Пусть $z \in \overline{x} \cap \overline{y}$. Тогда из предложения 2.1 следует, что $\overline{x} = \overline{z} = \overline{y}$. □

Предложение 2.3. Отношение эквивалентности индуцирует разбиение на множестве M, при этом множество M разбивается на непересекающиеся классы эквивалентности.

Верно и обратное. Всякое разбиение множества M индуцирует отношение эквивалентности на M.

Доказательство. Пусть на множестве M задано некоторое отношение эквивалентности. Тогда $M = \bigcup_{x \in M} \overline{x}$. При этом из предложения 2.2 следует, что каждый элемент $x \in M$ принадлежит одному и только одному классу эквивалентности: $x \in \overline{x}$.

Для доказательства второй части предложения можно положить $x \sim y$ тогда и только тогда, когда x и y принадлежат одному классу разбиения. □

Из сказанного выше следует такое

Предложение 2.4. Между всеми отношениями эквивалентности на множестве M и всеми разбиениями данного множества на непересекающиеся классы существует взаимно однозначное соответствие.

2.1.2. Отношение частичного порядка

Еще одним важным классом бинарных отношений являются отношения, обладающие свойствами рефлексивности, транзитивности и антисимметричности. Такие бинарные отношения называются *отношениями частичной упорядоченности*. Множество M с заданной на нем частичной упорядоченностью называется *частично упорядоченным* и обозначается $(M, \leqslant)$. Если $a \leqslant b$, то иногда говорят, что a предшествует b.

Пример 2.2. 1. Пусть M — множество всех действительных функций, определенных на некотором отрезке $[a, b]$. Для элементов $f, g \in M$ определим $f \leqslant g$ тогда и только тогда, когда для любого $x \in [a, b]$ выполнено неравенство $f(x) \leqslant g(x)$.

2. На множестве натуральных чисел можно определить частичную упорядоченность (отличную от обычного порядка), если понимать $x \leqslant y$ в том плане, что y делится без остатка на x.

3. Пусть M — некоторое множество. Обозначим через $\widetilde{M}$ множество всех подмножеств в M. Определим частичную упорядоченность на $\widetilde{M}$, полагая $A \leqslant B$ тогда и только тогда, когда $A \subseteq B$.

Если $a \leqslant b$ и $a \neq b$, то будем писать, что $a < b$ и говорить, что a меньше b. Бинарное отношение $<$ уже не является рефлексивным.

Пусть в множестве M задана частичная упорядоченность. Элементы a и b этого множества называются *сравнимыми*, если $a \leqslant b$ или $b \leqslant a$. Частично упорядоченное множество, в котором любые два элемента сравнимы, называется *линейно упорядоченным множеством* или *цепью*.

Пусть $(M, \leqslant)$, $(N, \leqslant)$ — два частично упорядоченных множества. Если существует такое взаимно однозначное соответствие $\varphi : M \to N$, что из $a \leqslant b$, $a, b \in M$, следует $\varphi(a) \leqslant \varphi(b)$ и обратно, то частично упорядоченные множества $(M, \leqslant)$ и $(N, \leqslant)$ называются *изоморфными*. Будем говорить, что частично упорядоченное множество $(M, \leqslant)$ изоморфно вкладывается в частично упорядоченное множество $(N, \leqslant)$, если существует такое подмножество X в N, что $(M, \leqslant)$ и $(X, \leqslant)$ изоморфны.

О важности третьего из указанных в 2.2 примеров говорит следующее утверждение.

Предложение 2.5. Всякое частично упорядоченное множество M изоморфно вкладывается в множество $\widetilde{N}$ всех подмножеств некоторого множества N, частично упорядоченного по включению. В качестве N можно взять, например, само M.

Доказательство. Пусть $a \in M$. Обозначим:

$$K_a = \{x \in M \mid x \leqslant a\}, \quad X = \{K_a \mid a \in M\} \subseteq \widetilde{M},$$

где $\widetilde{M}$ — множество всех подмножеств в M. Построим отображение $\varphi : M \to X$, поставив в соответствие каждому элементу

$a \in M$ множество $K_a \in X$. Пусть $K_a = K_b$ для некоторых $a, b \in M$. Тогда $a \leqslant b$, так как $a \in K_b$, и $b \leqslant a$, так как $b \in K_a$. Поэтому $a = b$ и отображение φ является инъекцией. В силу построения множества X отображение φ также является сюръекцией, поэтому φ — биекция.

Пусть $a \leqslant b$. Так как для любого $x \in K_a$ выполнено неравенство $x \leqslant a$, то $x \leqslant b$ и $K_a \subseteq K_b$. Обратно, если $K_a \subseteq K_b$, то $a \in K_b$ и $a \leqslant b$. Поэтому φ — изоморфное вложение M в $\widetilde{M}$. $\square$

Элемент a частично упорядоченного множества M называется *минимальным элементом* этого множества, если из неравенства $x \leqslant a$, $x \in M$, следует равенство $a = x$ (иными словами, в M нет ни одного элемента x, удовлетворяющего условию $x < a$).

Пример 2.3. 1. Частично упорядоченное множество $\widetilde{M}$ из примера 2.2 обладает единственным минимальным элементом — пустое подмножество.

2. В множестве всех непустых подмножеств множества M минимальными элементами являются все одноэлементные подмножества.

3. Пусть множество M бесконечно. Множество всех его бесконечных подмножеств не имеет минимальных элементов.

Предложение 2.6. Для частично упорядоченного множества $(M, \leqslant)$ следующие условия эквивалентны.

1. **Условие минимальности**. Всякое непустое подмножество N в M обладает хотя бы одним минимальным элементом.

2. **Условие индуктивности**. Пусть все минимальные элементы множества M обладают некоторым свойством Q и пусть из справедливости свойства Q для всех элементов, строго предшествующих некоторому элементу $a \in M$, может быть выведена справедливость этого свойства для самого элемента a. Тогда свойством Q обладают все элементы множества M.

3. **Условие обрыва убывающих цепей**. Всякая строго убывающая цепь элементов множества M :

$$a_1 > a_2 > \ldots > a_n > \ldots,$$

обрывается на конечном месте. Иными словами, для всякой убывающей цепи элементов:

$$a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n \geqslant \ldots$$

существует такой номер n, на котором эта цепь стабилизируется, т.е.:

$$a_n = a_{n+1} = \ldots .$$

Доказательство. 1 $\Rightarrow$ 2. Пусть выполнено условие минимальности и пусть в нем для некоторого свойства Q выполнены посылки условия индуктивности. Обозначим через N множество всех элементов в M, для которых свойство Q не выполнено. Предположим, что N не пусто. Из условия минимальности следует существование в N минимальных элементов. Пусть a — один из таких минимальных элементов. Данный элемент не является минимальным во всем множестве M, что следует из первой посылки условия индуктивности. Поэтому в M имеются элементы, строго предшествующие элементу a, причем все эти элементы обладают свойством Q. По второй посылке условия индуктивности и сам элемент a должен обладать свойством Q. Противоречие.

2 $\Rightarrow$ 3. Пусть выполнено условие индуктивности. Будем говорить, что элемент $a \in M$ обладает свойством Q, если всякая строго убывающая цепь элементов, начинающаяся от элемента a, обрывается на конечном месте. Понятно, что свойством Q обладают все минимальные элементы множества M. Пусть все элементы, предшествующие элементу a, обладают свойством Q. Тогда второй член любой строго убывающей цепи, начинающийся от элемента a, обладает свойством Q, поэтому данные цепи должны обрываться. Это означает, что элемент a также обладает свойством Q. Из условия индуктивности следует, что свойством Q обладают все элементы множества M.

3 $\Rightarrow$ 1. Пусть выполнено условие обрыва убывающих цепей. Предположим, что условие минимальности не выполнено. Пусть N — непустое подмножество в M, в котором нет минимальных элементов. Зафиксируем произвольный элемент

$a_1 \in N$. Так как в N нет минимальных элементов, то найдется такой элемент $a_2 \in N$, для которого выполнено неравенство $a_1 > a_2$ (в противном случае a_1 был бы минимальным элементом множества N). Пусть на n-м шаге построена убывающая конечная цепь $a_1 > a_2 > \ldots > a_n$. В N найдется элемент a_{n+1}, для которого $a_n > a_{n+1}$ (иначе a_n был бы минимальным элементом множества N). И так далее. Этот процесс будет бесконечным. Поэтому условие обрыва убывающих цепей не выполнено. Противоречие. □

Линейно упорядоченное множество, удовлетворяющее условию минимальности, называется *вполне упорядоченным*. Например, множество натуральных чисел с его естественным порядком является вполне упорядоченным. Из данного определения следуют такие несложные утверждения.

Предложение 2.7. 1. Всякое непустое подмножество вполне упорядоченного множества само вполне упорядочено.

2. Вполне упорядоченное множество обладает единственным минимальным элементом.

Пример 2.4. 1. Всякое конечное линейно упорядоченное множество является вполне упорядоченным.

2. Множество натуральных чисел с естественным порядком является вполне упорядоченным.

Трансфинитная индукция — метод доказательства, обобщающий математическую индукцию на случай несчетного числа значений параметра. Трансфинитная индукция основана на следующем утверждении. Пусть M — вполне упорядоченное множество, Q — свойство. Пусть:

1) минимальный элемент множества M обладает свойством Q;

2) из справедливости свойства Q для всех элементов, строго предшествующих некоторому элементу $a \in M$, следует справедливость этого свойства для самого элемента a.

Тогда все элементы множества M обладают свойством Q. Математическая индукция является частным случаем трансфи-

нитной индукции. Рассмотрим пример доказательства с помощью трансфинитной индукции.

Пример 2.5. Пусть множество $(M, \leqslant)$ вполне упорядочено и отображение $f : M \to M$ монотонно, т.е. из $x < y$, $x, y \in M$, следует $f(x) < f(y)$. Тогда $x \leqslant f(x)$ для любого $x \in M$.

Доказательство. Используя предложение 2.6, приведем три доказательства этого простого утверждения, чтобы поупражняться в применении метода трансфинитной индукции.

I. Понятно, что для минимального элемента множества M утверждение верно. Пусть N — множество всех таких элементов, для которых утверждение не верно. В N найдется минимальный элемент a, для которого $a > f(a)$. Пусть $b = f(a)$. Учитывая монотонность функции f, из неравенства $b < a$ следует $f(b) < f(a) = b$, т.е. $b > f(b)$. Но a — минимальный элемент множества N. Противоречие.

II. Применим условие индуктивности. Для минимального элемента посылка верна. Пусть $a \in M$ и для всех предшествующих к a элементов утверждение верно: для любого $x < a$ выполнено $x \leqslant f(x)$. Предположим, что для a утверждение не выполнено. Тогда $a > f(a) = b$. С одной стороны, $b \leqslant f(b)$ (посылка индукции). С другой стороны, используя монотонность f, $f(b) = f(f(a)) < f(a) = b$. Противоречие.

III. Предположим, что $a > f(a)$ для некоторого $a \in M$. Тогда по свойству монотонности получаем:

$$f(a) > f(f(a)), \quad f(f(a)) > f\big(f(f(a))\big)$$

и т.д. Это значит, что в M существует бесконечная убывающая цепь элементов:

$$a > f(a) > f(f(a)) > f\big(f(f(a))\big) > \dots,$$

что противоречит условию обрыва убывающих цепей. □

2.2. Алгебраические операции

Определение 2.1. *Внутренней (алгебраической) операцией* τ непустого множества X называется отображение $\tau : X^n \to X$,

при этом число n называется *арностью операции*. Внутренняя операция арности 2 называется бинарной операцией. Обычно бинарную операцию τ обозначают символами $+$, $\cdot$, $*$ и т.д., т.е. вместо $\tau(x, y)$ пишут, например, $x + y$, $x * y$, xy и т.д.

Пример 2.6. 1. Множество натуральных чисел $\mathbb{N}$ с бинарными операциями сложения и умножения. При этом заметим, что вычитание и деление не являются внутренними операциями на данном множестве.

2. Множество целых чисел $\mathbb{Z}$ с бинарными операциями сложения, вычитания и умножения.

3. Множество действительных чисел $\mathbb{R}$ с бинарными операциями сложения, вычитания и умножения. Заметим, что на всем $\mathbb{R}$ операция деления не является внутренней.

4. Множество $\mathbb{R} \setminus \{0\}$ с бинарными операциями умножения и деления. Теперь уже на множестве $\mathbb{R} \setminus \{0\}$ операции сложения и вычитания не будут внутренними.

5. Множество $M_n(\mathbb{R})$ всех матриц порядка n над $\mathbb{R}$ с бинарными операциями сложения, вычитания и умножения.

6. На множестве $\mathbb{R}$ можно определить бинарные операции min и max, которые возвращают соответственно минимальный и максимальный элементы для каждой пары.

7. Пусть $\widetilde{M}$ —множество всех подмножеств некоторого множества M. Операции пересечения и объединения являются бинарными операциями на множестве $\widetilde{M}$.

8. Пусть $F(M)$ — множество всех функций $f : M \to M$, т.е. множество всех преобразований множества M. Тогда операция композиции функций из $F(M)$ является бинарной операцией.

В дальнейшем бинарную операцию в общем случае будем записывать в мультипликативной форме: ab, xy.

Определение 2.2. Бинарная операция, заданная на множестве X, называется:

— *ассоциативной*, если для любых $x, y, z \in X$ выполнено равенство $(xy)z = x(yz)$;

— *коммутативной*, если для любых $x, y \in X$ выполнено равенство $xy = yx$.

Предложение 2.8. Если бинарная операция $*$, определенная на множестве X, является ассоциативной, то результат ее применения к n элементам не зависит от расстановки скобок, а зависит только от порядка расстановки данных элементов.

Доказательство. Для случая $n = 3$ утверждение верно по определению ассоциативной операции. Пусть утверждение верно для любого $k < n$. Пусть в выражении $a_1 a_2 \dots a_n$ произвольным образом расставлены скобки. Тогда выражение будет иметь вид:

$$(a_1 \dots a_s)(a_{s+1} \dots a_n),$$

где $1 \leqslant s \leqslant n-1$ и в выражениях $(a_1 \dots a_s)$ и $(a_{s+1} \dots a_n)$ также некоторым образом расставлены скобки. Заметим, что по предположению индукции расстановка скобок в двух последних выражениях не важна. Покажем, что:

$$(a_1 \dots a_s)(a_{s+1} \dots a_n) = (a_1 \dots a_{n-1})a_n.$$

Если $s = n-1$, то данное равенство верно. Пусть $s < n-1$. Обозначим $x = a_1 \dots a_s$, $y = a_{s+1} \dots a_{n-1}$. Требуемое равенство $x(ya_n) = (xy)a_n$ следует из определения ассоциативной операции. □

2.3. Полугруппы. Группы

Определение 2.3. Непустое множество G с заданной на нем бинарной операцией $*$ называется *полугруппой*, если операция $*$ является ассоциативной.

Пример 2.7. Множество $\mathbb{N}$ со следующими операциями является полугруппой:

1) $a * b = (a, b)$; 2) $a * b = [a, b]$; 3) $a * b = \min\{a, b\}$;
4) $a * b = a$; 5) $a * b = 1$.

Пример 2.8. Следующие подмножества множества всех действительных функций, определенных на множестве $\mathbb{R}$, относительно операции композиции являются полугруппами:

1) множество всех непрерывных функций;

2) множество всех ограниченных функций;

3) множество всех интегрируемых функций;

4) множество всех таких функций f, для которых $f(0) = 0$.

Пример 2.9. Пусть $\widetilde{M}$ — множество всех подмножеств непустого множества M. Тогда множество $\widetilde{M}$ относительно следующих операций является полугруппой:

1) пересечение;

2) объединение.

Определение 2.4. Непустое множество G с заданной на нем бинарной операцией $*$ называется *группой*, если относительно операции $*$ выполнены следующие условия:

1) операция ассоциативна;

2) относительно данной операции существует *нейтральный элемент* e такой, что $e * g = g * e = g$ для любого $g \in G$;

3) для любого $g \in G$ существует *симметричный элемент* g' такой, что $g' * g = g * g' = e$.

Группу G с операцией $*$ обозначают $(G, *)$. Если операция $*$ обладает свойством коммутативности, то группу называют *абелевой* (*коммутативной*).

Если в качестве бинарной операции выступает сложение, то группу $(G, +)$ называют *аддитивной*, нейтральный элемент называется нулевым, а симметричный элемент — противоположным. В случае же умножения группу $(G, \cdot)$ называют *мультипликативной*, нейтральный элемент называется единичным, а симметричный элемент — обратным.

В дальнейшем в качестве бинарной операции группы будем использовать мультипликативную форму записи.

Пример 2.10. 1. Аддитивные абелевы группы $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(M_n(\mathbb{R}), +)$.

2. Мультипликативные абелевы группы $(\mathbb{Q}\backslash\{0\}, \cdot)$, $(\mathbb{R}\backslash\{0\}, \cdot)$.

3. Пусть X — некоторое множество. Рассмотрим множество всех биективных отображений $f : X \to X$. Обозначим это множество через $S(X)$. Тогда относительно операции композиции функций $\circ$ получаем группу $(S(X), \circ)$.

4. Пусть $GL_n(\mathbb{R})$ — множество всех матриц из $M_n(\mathbb{R})$, которые имеют ненулевой определитель:

$$GL_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) \mid det(A) \neq 0\}.$$

Тогда множество $GL_n(\mathbb{R})$ с операцией умножения будет мультипликативной группой $(GL_n(\mathbb{R}), \cdot)$.

5. Пусть M — некоторое множество и $(G, *)$ — некоторая группа. Рассмотрим множество всех функций из множества M в множество G. Обозначим это множество через $F(M, G)$. Введем на множестве $F(M, G)$ бинарную операцию $\cdot$ следующим образом:

$$(fg)(x) = f(x) * g(x), \quad f, g \in F(M, G), \quad x \in M.$$

Получаем группу $(F(M, G), \cdot)$. Единичным элементом данной группы будет функция $\varepsilon : M \to G$, для которой $\varepsilon(x) = e$ для любого $x \in M$, где e — единичный элемент группы G. Обратной для функции $f \in F(M, G)$ будет функция f^{-1}, определенная таким образом: $f^{-1}(x) = (f(x))^{-1}$.

Предложение 2.9. (i) Всякая группа G обладает единственным единичным элементом.

(ii) Для всякого элемента g группы G существует единственный обратный элемент.

(iii) Если в группе G выполнено равенство $gg = g$ для некоторого $g \in G$, то $g = e$.

Доказательство. (i) Предположим что в некоторой группе G имеются два единичных элемента e_1 и e_2. Тогда:

$$e_1 = e_1 e_2 = e_2.$$

Первое равенство выполнено в силу того, что e_2 — единичный элемент, а второе — в силу того, что e_1 — тоже единичный элемент.

(ii) Пусть в некоторой группе G найдется такой элемент g, для которого имеются не менее двух обратных элементов g_1 и g_2. Тогда:

$$g_1 = g_1 e = g_1(gg_2) = (g_1 g)g_2 = eg_2 = g_2.$$

(*iii*) Домножив равенство $gg = g$ на g^{-1}, получаем, что $g = e$. □

Предложение 2.10. (*i*) Пусть G — некоторая группа и g — произвольный элемент группы G. Тогда отображения $\varphi_g,\ \psi_g : G \to G$, заданные правилами:

$$\varphi_g(x) = gx, \quad \psi_g(x) = g^{-1}xg,$$

являются биекциями.

(*ii*) Отображение $f : G \to G$, которое задается правилом $f(x) = x^{-1}$, является биекцией.

Доказательство. (*i*) Отображение φ_g является инъективным, так как если $\varphi_g(x) = \varphi_g(y)$ для некоторых $x, y \in G$, то, домножив слева равенство $gx = gy$ на g^{-1}, получим равенство $x = y$. Сюрьективность отображения φ_g следует из того, что для любого $y \in G$ выполнено равенство $\varphi_g(g^{-1}y) = y$.

Аналогично показывается биективность отображения ψ_g.

(*ii*) Справедливость данного пункта следует из предложения 2.9 и равенства $(x^{-1})^{-1} = x$. □

2.3.1. Подгруппы

Определение 2.5. Непустое подмножество H группы $(G, *)$ называется *подгруппой* группы G, если H относительно операции $*$, определенной в G, само является группой (обозначается $H \leqslant G$).

Заметим, что единичный элемент e_H подгруппы H всегда совпадает с единичным элементом e группы G. Действительно, так как e_H — единичный элемент в H, то выполнено равенство $e_H e_H = e_H$. Тогда из предложения 2.9 следует, что $e_H = e$.

Пример 2.11. 1. В любой группе G подгруппами являются такие множества: $H = \{e\}$, $H = G$. Данные подруппы называют тривиальными.

2. В абелевой аддитивной группе $(\mathbb{R}, +)$ подгруппами будут $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$.

3. В аддитивной группе $(M_n(\mathbb{R}), +)$ подгруппами будут $(UT_n(\mathbb{R}), +)$, $(Diag_n(\mathbb{R}), +)$, где $UT_n(\mathbb{R})$ — множество всех верхнетреугольных матриц в $M_n(\mathbb{R})$, $Diag_n(\mathbb{R})$ — множество всех диагональных матриц в $M_n(\mathbb{R})$.

4. В мультипликативной абелевой группе $(\mathbb{C} \setminus \{0\}, \cdot)$ подгруппой будет $(\{z \in \mathbb{C} \mid |z| = 1\}, \cdot)$.

5. Пусть $(G, *)$ — некоторая группа. Рассмотрим следующее подмножество в G :

$$C(G) = \{g \in G \mid gh = hg,\ h \in G\},$$

которое называется *центром группы* G. Тогда $(C(G), *)$ является подгруппой в G.

Предложение 2.11 (**эквивалентные условия подгруппы**)**.** Для произвольного непустого подмножества H группы G следующие условия эквивалентны:

(i) H является подгруппой;

(ii) для любого $g \in H$ обратный элемент g^{-1} также принадлежит H и для любой пары $g, h \in H$ элемент gh принадлежит H;

(iii) для любых $g, h \in H$ элемент $gh^{-1} \in H$.

Доказательство. Очевидно, что из (i) следует (ii). Обратно, пусть выполнено условие (ii). Так как H непусто, то в нем найдется хотя бы один элемент $g \in H$. Так как по условию g^{-1} принадлежит H, то по этому же условию (ii) единичный элемент $e = gg^{-1}$ тоже принадлежит H.

Очевидно, что из (ii) следует (iii). Обратно, пусть выполнено условие (iii). Так как H непусто, то найдется элемент $g \in H$. Тогда из условия (ii) следует, что $e = gg^{-1} \in H$. Пусть $h \in H$. Тогда $h^{-1} = eh^{-1} \in H$, т.е. множество H замкнуто относительно взятия обратного элемента.

Далее, пусть $g, h \in G$. Тогда $gh = g(h^{-1})^{-1} \in H$. Таким образом, из (iii) следует (ii). □

Предложение 2.12. Пересечение любой совокупности подгрупп в группе G будет являться подгруппой.

Пусть g — некоторый элемент группы G. Рассмотрим степени элемента g :

$$\dots g^{-3},\ g^{-2},\ g^{-1},\ g^0 = e,\ g,\ g^2,\ g^3, \dots, \tag{2.1}$$

где $g^n = \underbrace{g \dots g}_{n}$, $g^{-n} = \underbrace{g^{-1} \dots g^{-1}}_{n}$. Благодаря ассоциативности бинарной операции в группе, справедливы такие равенства:

$$g^m g^n = g^{m+n}, \quad (g^m)^n = g^{mn}, \quad m, n \in \mathbb{Z}.$$

Предложение 2.13. Элементы последовательности (2.1) образуют абелеву подгруппу в G.

Доказательство. Обозначим через H множество всех элементов из (2.1). Видно, что $e \in H$. Также для любых целых чисел m и n элемент $g^m g^n = g^{m+n}$ принадлежит H. Для любого целого n обратным к элементу g^n будет g^{-n}, который принадлежит H. Таким образом, учитывая предложение 2.11, получаем, что H является подгруппой. Коммутативность подгруппы H следует из равенств:

$$g^m g^n = g^{m+n} = g^{n+m} = g^n g^m, \quad m, n \in \mathbb{Z}. \qquad \square$$

Подгруппа, состоящая из элементов (2.1), называется *циклической подгруппой элемента* g и обозначается $\langle g \rangle$. Относительно последовательности (2.1) возможны случаи: 1) все элементы в (2.1) попарно различны; 2) найдутся такие целые числа m и n, $m < n$, что $g^m = g^n$. В первом случае элемент g называется *элементом бесконечного порядка.*

Рассмотрим второй случай. Из равенства $g^m = g^n$ следует равенство $g^{n-m} = e$, причем $n - m > 0$. Пусть d — минимальное из всех таких чисел $i = 1, \dots, n - m$, для которых $g^i = e$:

$$d = \min \{i = 1, \dots, n - m \mid g^i = e\}.$$

Тогда все элементы последовательности (2.1) будут принадлежать множеству:

$$H = \{g,\ g^2, \dots, g^d = e\}.$$

Действительно, пусть n — некоторое целое число. В силу единственности разложения $n = qd + r$, где $0 \leqslant r < d$, следует, что:

$$g^n = g^{qd+r} = (g^d)^q g^r = eg^r = g^r.$$

Очевидно также, что все элементы множества H попарно различны. Поэтому порядок циклической подгруппы H будет равен d. В этом случае элемент g называется *элементом конечного порядка* d, а число d — *порядком элемента* g.

Пример 2.12. Зафиксируем в группе $(\mathbb{Z}, +)$ некоторый элемент $m > 0$. Элементы:

$$\ldots -3m,\ -2m,\ -m,\ 0,\ m,\ 2m,\ 3m, \ldots$$

образуют циклическую подгруппу с образующим элементом m. При этом m имеет бесконечный порядок. Данная подгруппа обозначается $(m\mathbb{Z}, +)$.

Предложение 2.14. Любая подгруппа циклической группы $\langle g \rangle$ является циклической.

Доказательство. Если подгруппа H является тривиальной, то все очевидно. Поэтому пусть H — некоторая собственная подгруппа в $\langle g \rangle$. Обозначим через k наименьшее из положительных степеней элемента g, при условии, что $h = g^k \in H$. Тогда все элементы последовательности:

$$\ldots h^{-3},\ h^{-2},\ h^{-1},\ h^0 = e,\ h,\ h^2,\ h^3, \ldots,$$

исчерпывают множество H. Действительно, во-первых, все элементы данной последовательности принадлежат H, так как H является подгруппой. А во-вторых, если $g^n \in H$, то из разложения числа $n = kq + r$, $0 \leqslant r < k$, следует, что:

$$g^r = g^n (g^k)^{-q} \in H.$$

В силу выбора числа k получаем, что $r = 0$, поэтому $g^n = (g^k)^q$. $\square$

Предложение 2.15. Пусть $H = \{g, g^2, \ldots, g^n = e\}$ — конечная циклическая группа порядка n. Тогда элемент g^k, где

$1 \leqslant k \leqslant n-1$, является образующим элементом группы H тогда и только тогда, когда $(k, n) = 1$.

Доказательство. Пусть g^k является образующим группы H. Предположим, что $(k, n) = d > 1$. Тогда $k = d\widetilde{k}$, $n = d\widetilde{n}$ и:

$$(g^k)^{\widetilde{n}} = g^{\widetilde{k}d\widetilde{n}} = (g^n)^{\widetilde{k}} = e.$$

Так как $\widetilde{n} < n$, то g^k не является образующим группы H, противоречие.

Обратно, пусть число k, $1 \leqslant k \leqslant n-1$, взаимно просто с n. Рассмотрим последовательность:

$$g^k,\ (g^k)^2,\ \ldots, (g^k)^n.$$

Предположим, в данной последовательности найдутся два одинаковых значения: $(g^k)^i = (g^k)^j$, $1 \leqslant i < j \leqslant n$. Тогда $g^{k(j-i)} = e$, где $0 < j - i < n$. Так как $(k, n) = 1$ и n делит $k(j-i)$, то n делит $j - i$. С учетом того, что $j - i < n$ получаем, что $i = j$. Противоречие. Поэтому g^k является образующим группы H. □

Пример 2.13. Группа $(\mathbb{Z}_n, +)$ является циклической. Она порождается любым элементом $a \in \mathbb{Z}_n$ таким, что $(a, n) = 1$.

Предложение 2.16 (критерий конечной подгруппы). Произвольное непустое конечное подмножество H группы G является подгруппой тогда и только тогда, когда из $g, h \in H$ следует, что $gh \in H$.

Доказательство. Пусть $g \in H$. Тогда подгруппа $\langle g \rangle \subseteq H$ является конечной в силу конечности множества H и состоит из элементов g, g^2,..., $g^d = e$, где d — порядок элемента g. При этом $g^{-1} = g^{d-1} \in H$. Следовательно, по предложению 2.11 H является подгруппой. □

Предложение 2.17. Пусть G — некоторая абелева группа. Тогда множество всех ее элементов конечного порядка будет являться подгруппой в G.

Пусть M — некоторое подмножество группы G. Тогда существует минимальная подгруппа в G, содержащая множество M

(обозначается $\langle M \rangle$). Очевидно, что $\langle M \rangle$ является пересечением всех подгрупп в G, которые содержат множество M.

Предложение 2.18. Пусть M — подмножество в группе G. Тогда

$$\langle M \rangle = \{ g_1^{\varepsilon_1} \dots g_n^{\varepsilon_n} \mid g_i \in M,\ \varepsilon_i \in \{-1, 1\},\ n \in \mathbb{N} \}.$$

Если для некоторого множества $M \subseteq G$ выполнено равенство $\langle M \rangle = G$, то множество M называется *порождающим множеством группы* G.

Пример 2.14. 1. $\mathbb{Z} = \langle 1 \rangle$.
2. $\mathbb{Z}^* = \langle -1 \rangle$.

2.3.2. Смежные классы. Теорема Лагранжа

Пусть H — некоторая подгруппа группы G. Для произвольного элемента $g \in G$ обозначим:

$$gH = \{ gh \mid h \in H \}.$$

Множество gH называется *левым смежным классом группы* G *по подгруппе* H, определяемым элементом g. Так как $e \in H$, то $g \in gH$. Аналогично, множество Hg называется *правым смежным классом.*

Определим на множестве G отношение эквивалентности следующим образом:

$$a \sim b \Leftrightarrow a^{-1}b \in H, \quad a, b \in G.$$

Тогда для любого $a \in G$ класс эквивалентности, порожденный элементом a, в точности совпадет с левым смежным классом $aH : \overline{a} = aH$.

Аналогично, можно определить на множестве G такое отношение эквивалентности:

$$a \sim b \Leftrightarrow ab^{-1} \in H, \quad a, b \in G.$$

В этом случае для любого $a \in G$ верно равенство: $\overline{a} = Ha$.

Предложение 2.19. (i) Любой смежный класс порождается любым своим представителем, т.е. для любого $a \in gH$ $(a \in Hg)$ выполнено равенство $gH = aH$ $(Hg = Ha)$.

(ii) Любые два левых (правых) смежных класса либо не пересекаются, либо совпадают.

(iii) Между множеством левых смежных классов $\{gH \mid g \in G\}$ и множеством правых смежных классов $\{Hg \mid g \in G\}$ существует взаимно однозначное соответствие.

(iv) Для любого $g \in G$ множества H, gH и Hg являются равномощными.

Доказательство. Пункты (i) и (ii) следуют из предложений 2.1 и 2.2.

(iii) Для начала заметим, что из предложения 2.10 следуют такие равенства:

$$\{gH \mid g \in G\} = \{g^{-1}H \mid g \in G\},$$

$$\{Hg \mid g \in G\} = \{Hg^{-1} \mid g \in G\}.$$

Рассмотрим отображение $\psi : G \to G$, задаваемое правилом $\psi(x) = x^{-1}$. Тогда:

$$\psi(gH) = (gH)^{-1} = H^{-1}g^{-1} = Hg^{-1}.$$

(iv) Данный пункт следует из предложения 2.10. □

Из данного предложения следует, что группа G разбивается на непересекающиеся левые (правые) смежные классы по подгруппе H. Мощность множества смежных классов называется *индексом подгруппы H в группе G* и обозначается $G : H$. Из предложения 2.19 следует теорема 2.1.

Теорема 2.1 (Лагранж). Если группа G конечна, то для любой подгруппы H в G верно равенство:

$$|G| = |H| \cdot |G : H|.$$

Следствие 2.1. (i) Порядок любого элемента g конечной группы G является делителем порядка группы G, в частности, $g^{|G|} = e$.

(*ii*) Если порядок группы G является простым числом, то G является циклической группой, причем для любого $g \in G$, $g \neq e$, $G = \langle g \rangle$.

Алгоритм 2.1 (**определение порядка элемента группы при известной факторизации порядка n группы**). Порядок элемента конечной группы G порядка n можно определить по следующему алгоритму. Пусть $g \in G$, $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ — каноническое разложение числа n. Выходные данные: порядок d элемента g.

1. $d := n$.
2. Цикл: для i от 1 до k :
   2.1. $d := d/p_i^{\alpha_i}$.
   2.2. $b := g^d$.
   2.3. Цикл: пока $b \neq 1$:
      2.3.1. $b := b^{p_i}$.
      2.3.2. $d := d \cdot p_i$.
3. Вернуть d.

Алгоритм 2.2 (**поиск образующего элемента циклической группы**). Приведем вероятностный алгоритм поиска образующего элемента циклической группы порядка n. Эффективность алгоритма определяется тем, что группа содержит $\varphi(n)$ образующих элементов, и вероятность того, что случайно выбранный элемент является образующим равна $\dfrac{\varphi(n)}{n}$, причем $\dfrac{\varphi(n)}{n} > \dfrac{1}{6 \ln \ln n}$ (см. [33]). Пусть $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ — каноническое разложение числа n. Выходные данные: образующий элемент g группы G.

1. Выбрать случайный элемент g группы G.
2. Цикл: для i от 1 до k :
   2.1. $b := g^{n/p_i}$.
   2.2. Если $b = 1$, то перейти к 1.
3. Вернуть g.

2.3.3. Нормальная подгруппа

Определение 2.6. Подгруппа H группы G называется *нормальной подгруппой* (обозначается $H \trianglelefteq G$), если для любого элемента $g \in G$ выполнено равенство $gH = Hg$, т.е. левый смежный класс группы G по подгруппе H, определяемый элементом g, равен правому смежному классу.

Очевидно, что в абелевой группе любая подгруппа является нормальной подгруппой. Рассмотрим эквивалентные условия нормальной подгруппы.

Элементы a и b группы G называются *сопряженными* в G, если найдется такой элемент $g \in G$, что $b = g^{-1}ag$.

Предложение 2.20. Для любой подгруппы H группы G следующие условия эквивалентны:

(i) H является нормальной подгруппой;

(ii) для любого $g \in G$ выполнено равенство $g^{-1}Hg = H$;

(iii) подгруппа H с каждым своим элементом содержит все сопряженные с ним элементы, т.е. $g^{-1}Hg \subseteq H$, $g \in G$.

Доказательство. Эквивалентность пунктов (i) и (ii) следует из предложения 2.10.

Очевидно, что из (ii) следует (iii).

Обратно, пусть для любого $g \in G$ выполнено включение $g^{-1}Hg \subseteq H$. Покажем, что тогда будет выполнено и такое включение: $H \subseteq g^{-1}Hg$. Пусть $h \in H$. Тогда:

$$h = ehe = g^{-1}ghg^{-1}g = g^{-1}(g^{-1})^{-1}hg^{-1}g = g^{-1}\widetilde{h}g,$$

где $\widetilde{h} = (g^{-1})^{-1}hg^{-1}$. Так как $g^{-1}Hg \subseteq H$ для любого $g \in G$, то $\widetilde{h} \in H$. Поэтому $h = g^{-1}\widetilde{h}g \in g^{-1}Hg$. Таким образом, из (iii) следует (ii). □

2.3.4. Фактор-группа

Пусть H — нормальная подгруппа группы G. Рассмотрим смежные классы группы G по H (нет необходимости оговаривать левые или правые смежные классы, так как, в силу нормальности подгруппы H, они совпадают). Введем на множестве

всех смежных классов $\{gH \mid g \in G\}$ бинарную операцию:

$$aHbH = abH. \tag{2.2}$$

Покажем корректность данной операции. Пусть $\widetilde{a} \in aH$, $\widetilde{b} \in bH$, т.е. $\widetilde{a} = ah_1$, $\widetilde{b} = bh_2$ для некоторых $h_1, h_2 \in H$. Заметим, что в силу равенства $bH = Hb$ следует, что для некоторого $h_3 \in H$ выполнено равенство $h_1 b = bh_3$. Поэтому:

$$\widetilde{a}H\widetilde{b}H = \widetilde{a}\widetilde{b}H = ah_1bh_2H = abh_3h_2H = abH.$$

Предложение 2.21. Смежные классы группы G по нормальной подгруппе H с бинарной операцией (2.2) образуют группу, которая обозначается G/H.

Доказательство. В силу ассоциативности бинарной операции в группе G, операция (2.2) будет также ассоциативной:

$$(aHbH)cH = abHcH = (ab)cH = a(bc)H = aH(bHcH).$$

Единичным элементом относительно операции (2.2) будет сама подгруппа H :

$$(gH)H = gHeH = geH = gH = egH = eHgH = H(gH).$$

Обратным к элементу gH будет элемент $g^{-1}H$:

$$gHg^{-1}H = gg^{-1}H = eH = H = eH = g^{-1}gH = g^{-1}HgH. \quad \square$$

2.3.5. Морфизмы групп

Определение 2.7. Отображение φ из группы $(G, *)$ в группу $(\widetilde{G}, \circ)$ называется *гомоморфизмом*, если $\varphi(g * h) = \varphi(g) \circ \varphi(h)$ для любых $g, h \in G$. Очень часто употребляют мультипликативную запись $\varphi(gh) = \varphi(g)\varphi(h)$, при этом полагая, что в группах G и $\widetilde{G}$, в общем случае, различные бинарные операции.

Пример 2.15. 1. Множество $\{-1, 1\}$ с обычным умножением является мультипликативной абелевой группой $(\{-1, 1\}, \cdot)$. Отображение $sgn : \mathbb{R} \setminus \{0\} \to \{-1, 1\}$, сопоставляющее каждому действительному числу его знак, является гомоморфизмом группы $(\mathbb{R} \setminus \{0\}, \cdot)$ в $(\{-1, 1\}, \cdot)$.

2. Отображение группы $(\mathbb{Z}, +)$ в $(\{-1, 1\}, \cdot)$, определенное по правилу $\varphi(x) = (-1)^x$, является гомоморфизмом.

3. Отображение $\varphi(x) = |x|$ группы $(\mathbb{R} \setminus \{0\}, \cdot)$ в $(\mathbb{R}^+, \cdot)$ является гомоморфизмом.

4. Отображение $det : GL_n(\mathbb{R}) \to \mathbb{R}\setminus\{0\}$, сопоставляющее матрице ее определитель, будет гомоморфизмом группы $(GL_n(\mathbb{R}), \cdot)$ в $(\mathbb{R} \setminus \{0\}, \cdot)$.

Определение 2.8. *Ядром* гомоморфизма φ называется множество:

$$Ker\ \varphi = \{g \in G \mid \varphi(g) = \varepsilon\},$$

где ε — единичный элемент группы $\widetilde{G}$. *Образом* гомоморфизма φ называется множество:

$$Im\ \varphi = \{\varphi(g) \mid g \in G\}.$$

Предложение 2.22. Пусть $\varphi : G \to \widetilde{G}$ — гомоморфизм. Тогда:

(i) $\varphi(e) = \varepsilon$, где e, ε — единичные элементы соответствующих групп G и $\widetilde{G}$, и для любого $g \in G$ выполнено $\varphi(g^{-1}) = (\varphi(g))^{-1}$;

(ii) $Ker\ \varphi \trianglelefteq G$, $Im\ \varphi \leqslant \widetilde{G}$.

Доказательство. (i) Из равенства $\varphi(e)\varphi(e) = \varphi(e)$ следует, что $\varphi(e) = \varepsilon$ (предложение 2.9). Далее, пусть $g \in G$. Тогда:

$$\varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = \varphi(e) = \varepsilon =$$
$$= \varphi(e) = \varphi(g^{-1}g) = \varphi(g^{-1})\varphi(g).$$

Поэтому $(\varphi(g))^{-1} = \varphi(g^{-1})$.

(ii) Покажем, что множество $Ker\ \varphi$ замкнуто относительно произведения элементов и взятия обратного элемента. Пусть $g, h \in Ker\ \varphi$. Тогда:

$$\varphi(gh) = \varphi(g)\varphi(h) = \varepsilon\varepsilon = \varepsilon,$$
$$\varphi(g^{-1}) = (\varphi(g))^{-1} = \varepsilon^{-1} = \varepsilon.$$

Итак, $Ker\ \varphi$ является подгруппой в G (предложение 2.11). Покажем, что $Ker\ \varphi$ — нормальная подгруппа. Пусть $h \in Ker\ \varphi$, $g \in G$. Тогда:

$$\varphi(g^{-1}hg) = \varphi(g^{-1})\varphi(h)\varphi(g) = \varphi(g^{-1})\varepsilon\varphi(g) =$$

$$= \varphi(g^{-1})\varphi(g) = \varphi(g^{-1}g) = \varphi(e) = \varepsilon.$$

Поэтому $g^{-1}hg \in Ker\ \varphi$, из чего следует, что:

$$g^{-1}(Ker\ \varphi)g \subseteq Ker\ \varphi.$$

Таким образом, из предложения 2.20 следует, что $Ker\ \varphi$ — нормальная подгруппа в G.

Замкнутость множества $Im\ \varphi$ относительно произведения элементов и взятия обратного элемента следует из равенств:

$$\varphi(g)\varphi(h) = \varphi(gh), \quad (\varphi(g))^{-1} = \varphi(g^{-1}),$$

поэтому $Im\ \varphi$ является подгруппой в $\widetilde{G}$. □

Определение 2.9. Группы G и $\widetilde{G}$ называются *изоморфными* (обозначается $G \cong \widetilde{G}$), если существует биективное отображение $\varphi : G \to \widetilde{G}$, которое является гомоморфизмом групп.

Пример 2.16. 1. Группы $(\mathbb{Z}, +)$ и $(2\mathbb{Z}, +)$ изоморфны, при этом отображение $\varphi : \mathbb{Z} \to 2\mathbb{Z}$, которое определяется правилом $\varphi(z) = 2z$, $z \in \mathbb{Z}$, является изоморфизмом групп.

2. Пусть $\mathbb{R}^+$ — множество всех положительных действительных чисел. Данное множество с операцией умножения является мультипликативной абелевой группой $(\mathbb{R}^+, \cdot)$. Данная группа изоморфна аддитивной группе $(\mathbb{R}, +)$. В качестве изоморфного отображения можно рассмотреть отображение $\varphi : \mathbb{R}^+ \to \mathbb{R}$, определенное правилом $\varphi(x) = ln\ x$.

Теорема 2.2 (о гомоморфизме групп). Пусть $\varphi : G \to \widetilde{G}$ — гомоморфизм. Тогда $G/Ker\ \varphi \cong Im\ \varphi$.

Доказательство. Определим отображение:

$$f : G/Ker\ \varphi \to Im\ \varphi$$

следующим образом:

$$f(gKer\ \varphi) = \varphi(g), \quad g \in G.$$

Проверим корректность отображения f. Пусть $h \in gKer\ \varphi$. Тогда $h = gk$ для некоторого $k \in Ker\ \varphi$. Поэтому:

$$f(hKer\ \varphi) = \varphi(h) = \varphi(gk) = \varphi(g)\varphi(k) =$$

$$= \varphi(g)\varepsilon = \varphi(g) = f(gKer\ \varphi).$$

Далее, отображение f является гомоморфизмом, так как:

$$f(gKer\varphi\ hKer\varphi) = f(ghKer\varphi)) = \varphi(gh) =$$

$$= \varphi(g)\varphi(h) = f(gKer\ \varphi)f(hKer\ \varphi).$$

Осталось показать, что f является биекцией. Очевидно, что f — сюръекция. Покажем, что f еще и инъективное отображение. Пусть:

$$\varphi(g) = f(gKer\ \varphi) = f(hKer\ \varphi) = \varphi(h).$$

Тогда:

$$\varepsilon = \varphi(g)(\varphi(h))^{-1} = \varphi(g)\varphi(h^{-1}) = \varphi(gh^{-1}).$$

Поэтому $gh^{-1} \in Ker\ \varphi$ и $gKer\ \varphi = hKer\ \varphi$. □

Теорема 2.3 (**Кэли**). Любая группа $(G, *)$ изоморфно вкладывается в группу подстановок $(S(G), \circ)$ множества G. В частности, любая конечная группа порядка n изоморфно вкладывается в симметрическую группу S_n.

Доказательство. Пусть $g \in G$. Обозначим через φ_g отображение $\varphi_g : G \to G$, заданное правилом $\varphi_g(x) = gx$, $x \in G$. В предложении 2.10 показано, что φ_g является биекцией, поэтому $\varphi_g \in S(G)$. При этом для любых $g, h \in G$ выполнено равенство $\varphi_g \circ \varphi_h = \varphi_{gh}$, так как:

$$(\varphi_g \circ \varphi_h)(x) = \varphi_g(\varphi_h(x)) = \varphi_g(hx) = g(hx) = (gh)x = \varphi_{gh}(x)$$

Определим отображение $f : G \to S(G)$ следующим образом: $f(g) = \varphi_g$. Для доказательства теоремы достаточно показать, что f является инъективным гомоморфизмом. Равенство:

$$f(gh) = \varphi_{gh} = \varphi_g \circ \varphi_h = f(g) \circ f(h).$$

показывает, то f — гомоморфизм. Предположим, что для некоторых $g, h \in G$ выполнено $f(g) = f(h)$. Тогда $\varphi_g = \varphi_h$, т.е. для любого $x \in G$ выполнено равенство $gx = hx$. В частности, $g = ge = he = h$. Поэтому f — инъекция. □

2.4. Кольца

Определение 2.10. *Кольцом* называется множество R с двумя бинарными операциями $+$ и $\cdot$ (обозначается $(R, +, \cdot)$), если

1. $(R, +)$ является аддитивной абелевой группой с нулевым элементом 0.

2. Операции $+$ и $\cdot$ связаны законом дистрибутивности:

$$(a+b)c = ac + bc, \quad a(b+c) = ab + ac, \quad a, b, c \in R.$$

Противоположный к элементу a по сложению обозначается $-a$.

Предложение 2.23. Для любых элементов a, b, c произвольного кольца R верны следующие утверждения:

(i) $a0 = 0a = 0$;

(ii) $-(-a) = a$;

(iii) $-(ab) = (-a)b = a(-b)$;

(iv) $(-a)(-b) = ab$;

(v) законы дистрибутивности выполняются и для разности:

$$(a-b)c = ac - bc, \quad a(b-c) = ab - ac.$$

Доказательство. (i) К обеим частям равенства $a0 = a0 + a0$, которое получается из:

$$a0 = a(0+0) = a0 + a0,$$

прибавим $-a0$. Поучаем $a0 = 0$. Аналогично, $0a = 0$.

(ii) Так как элемент $(-a)$ является противоположным к a, то из равенств $a + (-a) = (-a) + a = 0$ следует, что противоположным к элементу $(-a)$ является a, поэтому $-(-a) = a$.

(iii) $ab + (-a)b = (a + (-a))b = 0b = 0$.

(iv) Следует из пунктов (ii) и (iii).

(v) В силу ассоциативности операции сложения, имеет место равенство:

$$(a-b) + b = a.$$

Домножим справа данное равенство на c и применим закон дистрибутивности:

$$(a-b)c + bc = ac.$$

К обеим частям последнего равенства осталось прибавить элемент $-bc$. $\square$

Следствие 2.2. Во всяком кольце с единицей R, в котором не менее двух элементов, $0 \neq 1$.

Доказательство. Пусть R — некоторое кольцо с единицей, содержащее не менее двух элементов. Предположим, что $0 = 1$. Пусть $a \in R$. Тогда $a = a1 = a0 = 0$ (предложение 2.23), т.е. $R = \{0\}$. Противоречие. $\square$

Определение 2.11. Кольцо R называется:

— *ассоциативным*, если операция $\cdot$ обладает свойством ассоциативности: $(ab)c = a(bc)$, $a, b, c \in R$;

— *коммутативным*, если операция $\cdot$ обладает свойством коммутативности: $ab = ba$, $a, b \in R$;

— *с единицей* 1, если 1 является единичным элементом относительно операции умножения: $a1 = 1a = a$ для любого $a \in R$.

В дальнейшем будем рассматривать только ассоциативные кольца и под понятием кольца будем подразумевать именно такие кольца.

Пример 2.17. 1. Пусть $(G, +)$ — аддитивная абелева группа. Введем на множестве G бинарную операцию $\cdot$, определив $ab = 0$ для любых $a, b \in G$. Получаем коммутативное кольцо $(G, +, \cdot)$. Оно называется кольцом с нулевым умножением.

2. Коммутативные кольца с единицей $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$.

3. Коммутативное кольцо (без единицы) $(2\mathbb{Z}, +, \cdot)$.

4. Кольца с единицей $(M_n(\mathbb{R}), +, \cdot)$, $(UT_n(\mathbb{R}), +, \cdot)$, $(Diag_n(\mathbb{R}), +, \cdot)$. Заметим, что множество невырожденных матриц $GL_n(\mathbb{R})$ кольцом не является, так как операция сложения в $GL_n(\mathbb{R})$ не является замкнутой.

5. Рассмотрим множество упорядоченных пар действительных чисел $\mathbb{R}^2 = \{(a, b) \mid a, b \in \mathbb{R}\}$. Введем на данном множестве бинарные операции $+$ и $\cdot$ следующим образом:

$$(a, b) + (c, d) = (a + c, b + d),$$

$$(a, b)(c, d) = (ac, bd).$$

Получаем коммутативное кольцо с единицей $(\mathbb{R}^2, +, \cdot)$.

6. Пусть M — некоторое множество и $(R, +, \cdot)$ — некоторое кольцо. Рассмотрим множество всех функций из M в R. Обозначим данное множество через $F(M, R)$. Определим на множестве (F, R) операции $+$ и $\cdot$ следующим образом:

$$(f + g)(x) = f(x) + g(x),$$

$$(fg)(x) = f(x)g(x),$$

где $f, g \in F(M, R)$, $x \in M$. Получаем кольцо $(F(M, R), +, \cdot)$. Данное кольцо называется *полным кольцом функций из M в R*. Заметим, что свойства, которым обладало кольцо R (ассоциативность, коммутативность, наличие единицы), переносятся на кольцо $F(M, R)$.

Элемент a кольца R с единицей называется *обратимым*, если для него найдется такой элемент $a^{-1} \in R$, что $aa^{-1} = a^{-1}a = 1$.

Множество всех обратимых элементов кольца R обозначается через R^*. Например:

$$\mathbb{R}^* = \mathbb{R} \setminus \{0\}, \quad \mathbb{Z}^* = \{-1, 1\}, \quad (M_n(\mathbb{R}))^* = GL_n(\mathbb{R}).$$

Предложение 2.24. Пусть $(R, +, \cdot)$ — кольцо с единицей. Тогда $(R^*, \cdot)$ — мультипликативная группа.

Доказательство. Пусть $a \in R^*$. Тогда $a^{-1} \in R^*$, так как $(a^{-1})^{-1} = a \in R$. Также если $a, b \in R^*$, то $ab \in R^*$, так как $(ab)^{-1} = b^{-1}a^{-1}$, т.е. для ab имеется обратный элемент $b^{-1}a^{-1}$ кольца R. □

Группа $(R^*, \cdot)$ называется *мультипликативной группой* кольца R.

Ненулевые элементы a и b кольца R называются делителями нуля, если $ab = 0$. Например, в кольцах $\mathbb{Z}$, $2\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ нет делителей нуля. Если в кольце нет делителей нуля, то оно называется *кольцом без делителей нуля*. Коммутативное кольцо с $1 \neq 0$ без делителей нуля называют *целостным кольцом или областью целостности*.

Пример 2.18 (кольца с делителями нуля)**.** 1. В кольце матриц $M_n(\mathbb{R})$, $n > 1$. делителями нуля будут матричные единички e_{11} и e_{nn}.

2. В кольце $(\mathbb{R}^2, +, \cdot)$ из примера 2.17.5 делителями нуля будут $(1, 0)$ и $(0, 1)$.

3. Рассмотрим кольцо $(F(M, R), +, \cdot)$ из примера 2.17.6. Пусть каждое из множеств M и R содержит не менее двух элементов. Разобьем множество M на два непустых непересекающихся класса A и B. Зафиксируем также элемент $r \in R$, причем $r \neq 0$. Рассмотрим такие два отображения $f, g \in F(M, R)$, что:

$$f(a) = 0,\ a \in A, \quad f(b) = r,\ b \in B,$$

$$g(a) = r,\ a \in A, \quad f(b) = 0,\ b \in B.$$

Тогда элементы f и g будут являться делителями нуля кольца $(F(M, R), +, \cdot)$.

2.4.1. Кольца многочленов

Пусть R — некоторое кольцо. С помощью нового элемента $x \notin R$ образуем выражение вида $P_n = \sum_{k=0}^{n} a_k x^k$, где все $a_k \in R$. Такие выражения называются *многочленами (полиномами) над кольцом* R, x называется *переменной*, элементы a_k — *коэффициентами. Степенью многочлена* называется наибольшее значение k, при котором $a_k \neq 0$, и обозначается $\deg\ P_n$. Будем считать, что элемент x перестановочен со всеми элементами кольца R. Обозначим множество всех многочленов над R через $R[x]$.

Два многочлена из $R[x]$ называются *равными*, если равны их коэффициенты при одинаковых степенях переменной x.

Введем на множестве $R[x]$ операции сложения и умножения:

$$\sum_{k=0}^{m} a_k x^k + \sum_{k=0}^{n} b_k x^k = \sum_{k=0}^{\max\{m,n\}} (a_k + b_k) x^k, \tag{2.3}$$

$$\sum_{k=0}^{m} a_k x^k \cdot \sum_{k=0}^{n} b_k x^k = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i b_j \right) x^k. \tag{2.4}$$

Предложение 2.25. Если R — ассоциативное коммутативное кольцо с единицей, то множество $R[x]$ с бинарными операциями сложения и умножения является ассоциативным коммутативным кольцом с единицей.

Аналогично, к кольцу $R[x_1]$ можно присоединить новую переменную $x_2 \notin R[x_1]$. В силу перестановочности x_1 и x_2 получим кольцо от двух переменных $R[x_1][x_2] = R[x_1, x_2]$. Продолжая далее, можно получить кольцо от n переменных $R[x_1, \ldots, x_n]$, при этом многочлены из $R[x_1, \ldots, x_n]$ имеют вид:

$$\sum_{k_1 \geqslant 0, \ldots, k_n \geqslant 0} a_{k_1 \ldots k_n} x_1^{k_1} \ldots x_n^{k_n}$$

с конечным числом слагаемых. *Степенью многочлена от n переменных* называют наибольшую сумму $k_1 + \ldots + k_n$ при условии $a_{k_1 \ldots k_n} \neq 0$.

Предложение 2.26. Если кольцо R целостное, то для любого натурального n кольцо $R[x_1, \ldots, x_n]$ также является целостным.

Предложение 2.27. Пусть $A(x)$ и $B(x)$ — некоторые многочлены над полем K и $B(x)$ — ненулевой многочлен. Тогда существует единственное разложение вида

$$A(x) = Q(x)B(x) + R(x), \tag{2.5}$$

где $Q(x)$ и $R(x)$ — многочлены на полем K, причем выполнено неравенство $\deg\ R(x) < \deg\ B(x)$.

Доказательство. Пусть:

$$A(x) = a_0 + a_1 x + \ldots + a_n x^n, \quad a_n \neq 0, \quad \deg\ A(x) = n,$$

$$B(x) = b_0 + b_1 x + \ldots + b_m x^m, \quad b_m \neq 0, \quad \deg\ B(x) = m.$$

Если $n < m$, то $A(x) = 0 \cdot B(x) + A(x)$. Если $n = m$, то разность:

$$R(x) = A(x) - \frac{a_n}{b_n} B(x)$$

является многочленом над полем K степени строго меньшей чем n. Поэтому в этом случае $Q(x) = a_n / b_n$.

Индукцией по $n - m$ покажем справедливость разложения (2.5). База индукции при $n - m \leqslant 0$ проверена. Пусть $n > m$. Обозначим:

$$\widetilde{A}(x) = A(x) - \frac{a_n}{b_m} x^{n-m} B(x).$$

Тогда степень многочлена $\widetilde{A}(x)$ строго меньше чем n. Поэтому по предположению индукции найдутся такие $\widetilde{Q}(x), R(x) \in K[x]$, что:

$$\widetilde{A}(x) = \widetilde{Q}(x)B(x) + R(x), \quad \deg\ R(x) < m.$$

Поэтому:

$$A(x) = \left(\frac{a_n}{b_m} x^{n-m} + \widetilde{Q}(x)\right) B(x) + R(x),$$

что доказывает разложение (2.5).

Покажем единственность такого разложения. Пусть:

$$A(x) = Q_1(x)B(x) + R_1(x), \quad A(x) = Q_2(x)B(x) + R_2(x),$$

$$\deg\ R_i(x) < m, \quad i = 1, 2.$$

Тогда:

$$(Q_1(x) - Q_2(x))B(x) = R_2(x) - R_1(x).$$

Понятно, что $\deg\ (R_2(x) - R_1(x)) < m$. Поэтому данное равенство возможно лишь при $R_1(x) = R_2(x)$, из которого уже следует (в силу предложения 2.26), что $Q_1(x) = Q_2(x)$. □

2.4.2. Подкольца

Определение 2.12. Непустое подмножество S кольца R называется *подкольцом*, если S относительно операций сложения и умножения, определенных в R, само является кольцом (обозначается $S \leqslant R$).

Пример 2.19. 1. В кольце $(\mathbb{R}, +, \cdot)$ подкольцами будут $(\mathbb{Q}, +, \cdot)$, $(\mathbb{Z}, +, \cdot)$, $(m\mathbb{Z}, +, \cdot)$ для произвольного натурального m.

2. В кольце матриц $(M_n(\mathbb{R}), +, \cdot)$ подкольцами будут $(M_n(\mathbb{Q}), +, \cdot)$, $(UT_n(\mathbb{R}), +, \cdot)$, $(Diag_n(\mathbb{R}), +, \cdot)$.

Предложение 2.28 (**критерий подкольца**)**.** Непустое подмножество S кольца R является подкольцом тогда и только тогда, когда для любых $x, y \in S$ выполнено $x - y \in S$, $xy \in S$.

Доказательство. Так как S непусто, то найдется $x \in S$. Тогда $0 = x - x \in S$. Поэтому для любого $y \in S$ обратный по сложению элемент $-y = 0 - y \in S$. Пусть $x, y \in S$. Так как $-y \in S$ и $-(-y) = y$, то $x + y = x - (-y) \in S$. Следовательно $(S, +)$ является аддитивной абелевой группой.

Также в силу замкнутости операции умножения в S подмножество S является подкольцом. □

Если S — некоторое подкольцо произвольного кольца R, то нулевой элементы 0_S подкольца S совпадает с нулевым элементом 0 кольца R. Относительно единицы в кольцах с единицами это не так. Например, рассмотрим подкольцо S в кольце матриц порядка два $(M_2(\mathbb{R}), +, \cdot)$ следующего вида:

$$S = \left\{ \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \mid a \in \mathbb{R} \right\}.$$

Единицей в подкольце S будет являться матричная единичка e_{11}, не совпадающая с единицей кольца M_2.

Предложение 2.29. Пусть R — кольцо с единицей и без делителей нуля. Тогда единица любого ненулевого подкольца S в R совпадает с единицей кольца R.

Доказательство. Пусть e и e_S — соответственно единицы кольца R и подкольца S. Тогда выполнены равенства:

$$ee_S = e_S, \quad e_S e_S = e_S.$$

Поэтому $ee_S = e_S e_S$, из чего следует, что $(e - e_S)e_S = 0$. Так как в кольце R нет делителей нуля, то $e - e_S = 0$. □

Предложение 2.30. Пересечение любой совокупности подколец кольца R будет являться подкольцом.

Пусть M — некоторое подмножество кольца R. Обозначим через $\{M\}$ минимальное подкольцо в R, содержащее множество M. При этом $\{M\}$ является пересечением всех подколец в R, которые содержат множество M.

Предложение 2.31. Подкольцо $\{M\}$ состоит из всех конечных сумм вида:

$$\sum_i z_i a_{i_1} \dots a_{i_k}, \quad z_i \in \mathbb{Z},\ a_{i_j} \in M,\ k \in \mathbb{N}.$$

2.4.3. Идеалы кольца

Определение 2.13. Подкольцо I_l (I_r) кольца R называется *левым* (*правым*) *идеалом*, если для любых $i \in I_l$ ($i \in I_r$), $x \in R$ выполнено $xi \in I_l$ ($ix \in I_r$). Если идеал I является одновременно левым и правым, то его называют *двусторонним идеалом* и обозначают $I \trianglelefteq R$.

Например, в коммутативных кольцах все идеалы являются двусторонними. В дальнейшем двусторонние идеалы будем просто называть идеалами Понятие идеала является аналогом нормальной подгруппы в теории групп.

В любом кольце R идеалами будут нулевое подкольцо (состоящее из одного нуля) и само кольцо R. Данные идеалы называются *несобственными идеалами*. Все остальные идеалы в R носят название *собственных идеалов*.

Пример 2.20. 1. Пусть R — некоторое кольцо и $a \in R$. Обозначим $Ra = \{xa \mid x \in R\}$. Тогда множество Ra является левым идеалом в R. Аналогично, множество aR является правым идеалом в R.

2. В кольце целых чисел $\mathbb{Z}$ идеалами будут подкольца вида $m\mathbb{Z}$ для любого натурального m.

Предложение 2.32. (i) Сумма произвольных идеалов I_1 и I_2 кольца R является идеалом.

(ii) Пересечение произвольного семейства идеалов кольца R является идеалом.

2.4.4. Фактор-кольцо

Пусть I — некоторый идеал произвольного кольца R. Поскольку аддитивная группа $(I, +)$ является нормальной в груп-

пе $(R,+)$, то это позволяет определить аддитивную фактор-группу $(R/I,+)$ с операцией сложения:

$$(x+I)+(y+I)=(x+y)+I, \quad x,y\in R.$$

Превратим эту фактор-группу в фактор-кольцо $(R/I,+,\cdot)$, определив операцию умножения следующим образом:

$$(x+I)(y+I)=xy+I.$$

Проверим корректность данной операции. Пусть $a\in x+I$, $b\in y+I$. Тогда $a=x+i_1$, $b=y+i_2$ для некоторых $i_1,i_2\in I$. Имеем:

$$(a+I)(b+I)=ab+I=(x+i_1)(y+i_2)+I=(xy+xi_2+i_1y+i_1i_2)+I.$$

Так как I является идеалом, то $xi_2+i_1y+i_1i_2\in I$, поэтому:

$$(xy+xi_2+i_1y+i_1i_2)+I=xy+(xi_2+i_1y+i_1i_2)+I=$$

$$=xy+I=(x+I)(y+I).$$

Таким образом:

$$(a+I)(b+I)=(x+I)(y+I),$$

что показывает корректность операции умножения.

2.4.5. Кольцо классов вычетов

Пусть m — некоторое фиксированное натуральное число. Как было отмечено в примере 2.20, подкольцо $(m\mathbb{Z},+,\cdot)$ кольца целых чисел $(\mathbb{Z},+,\cdot)$ является идеалом. Рассмотрим фактор-кольцо $(\mathbb{Z}/m\mathbb{Z},+,\cdot)$, которое обозначается $\mathbb{Z}_m$, и называется кольцом вычетов по модулю m. Элементы в $\mathbb{Z}_m$ имеют вид $a+m\mathbb{Z}$, $a\in\mathbb{Z}$. Класс, порожденный элементом $a\in\mathbb{Z}$, будем обозначать через $\overline{a}$:

$$\overline{a}=a+m\mathbb{Z}=\{a+mt \mid t\in\mathbb{Z}\},$$

т.е. в классе $\overline{a}$ лежат все элементы из $\mathbb{Z}$, сравнимых с элементом a по модулю m :

$$x\in\overline{a}\Longleftrightarrow x\equiv a\ (\mathrm{mod}\ m)\Longleftrightarrow x-a\in m\mathbb{Z}.$$

При разбиении множества $\mathbb{Z}$ на классы эквивалентных элементов, множество $\mathbb{Z}_m$ будет состоять из следующих классов:

$$\mathbb{Z}_m = \{\overline{0},\ \overline{1},\ \ldots, \overline{m-1}\}.$$

Операции сложения и умножения в $\mathbb{Z}_m$ определены следующим образом:

$$\overline{a} + \overline{b} = \overline{a+b}, \quad \overline{a} \cdot \overline{b} = \overline{a \cdot b}.$$

Кольцо $\mathbb{Z}_m$ является ассоциативным коммутативным кольцом с единицей. В роли единицы выступает класс $\overline{1}$, а в роли нуля — $\overline{0}$.

Предложение 2.33. В кольце вычетов $\mathbb{Z}_m$ каждый элемент $\overline{a} \neq \overline{0}$ либо обратим, либо является делителем нуля, причем:

(i) $\overline{a}$ — обратим тогда и только тогда, когда $(a, m) = 1$;

(ii) $\overline{a}$ — делитель нуля тогда и только тогда, когда $(a, m) \neq 1$.

Доказательство. (i) Из теоремы 1.11 следует, что линейное диофантово уравнение $ax + my = 1$ имеет решение тогда и только тогда, когда $(a, m) = 1$. При этом:

$$\overline{1} = \overline{ax + my} = \overline{a} \cdot \overline{x} + \overline{m} \cdot \overline{y} = \overline{a} \cdot \overline{x} + \overline{0} = \overline{a} \cdot \overline{x}.$$

Поэтому $\overline{a}^{-1} = \overline{x}$.

(ii) Пусть $(a, m) = d > 1$. Тогда $a = sd$, $m = td$, $1 < t < m$ и:

$$\overline{a} \cdot \overline{t} = \overline{at} = \overline{sdt} = \overline{sm} = \overline{0}. \qquad \square$$

Следствие 2.3. (i) Кольцо $\mathbb{Z}_m$ является полем тогда и только тогда, когда m — простое число.

(ii) Порядок мультипликативной группы $\mathbb{Z}_m^*$ равен $\varphi(m)$, где φ — функция Эйлера.

2.4.6. Морфизмы колец

Определение 2.14. Пусть R и $\widetilde{R}$ — некоторые кольца. Отображение $\varphi : R \to \widetilde{R}$ называется *гомоморфизмом колец*, если:

$$\varphi(x + y) = \varphi(x) + \varphi(y), \quad \varphi(xy) = \varphi(x)\varphi(y), \quad x, y \in R.$$

Ядром гомоморфизма φ называется множество:

$$Ker\ \varphi = \{x \in R \mid \varphi(x) = 0\}.$$

Образом гомоморфизма φ называется множество:

$$Im\ \varphi = \{\varphi(x) \mid x \in R\}.$$

Предложение 2.34. Пусть $\varphi : R \to \widetilde{R}$ — гомоморфизм колец. Тогда:

(i) $\varphi(0) = 0$ и для любого $x \in R$ выполнено $\varphi(-x) = -\varphi(x)$;

(ii) $Ker\ \varphi \trianglelefteq R$, $Im\ \varphi \leqslant \widetilde{R}$.

Определение 2.15. Кольца R и $\widetilde{R}$ называются *изоморфными* (обозначается $R \cong \widetilde{R}$), если существует биективное отображение $\varphi : R \to \widetilde{R}$, которое является гомоморфизмом колец.

Теорема 2.4 (о гомоморфизме колец). Пусть $\varphi : R \to \widetilde{R}$ — гомоморфизм колец. Тогда $R/Ker\ \varphi \cong Im\ \varphi$.

Предложение 2.35. Всякое кольцо R изоморфно вкладывается в полное кольцо функций $F(M, R)$ (см. пример 2.17.6).

Доказательство. Обозначим через φ_r, $r \in R$, отображение из M в R, действующую по правилу: $\varphi_r(x) = r$ для любого $x \in M$. Тогда множество $S = \{\varphi_r \mid r \in R\}$ будет подкольцом в $F(M, R)$. Определим отображение $f : R \to S$ следующим образом: $f(r) = \varphi_r$. Тогда f является изоморфизмом колец R и S. □

2.4.7. Кольца главных идеалов

Определение 2.16. Идеал I коммутативного кольца R называется *главным*, если он имеет вид $I = aR$ для некоторого $a \in R$.

Предложение 2.36. В кольце целых чисел $\mathbb{Z}$ каждый идеал является главным.

Доказательство. Пусть I — некоторый идеал в $\mathbb{Z}$. Если идеал I нулевой, то $I = 0\mathbb{Z}$. Поэтому пусть $I \neq \{0\}$. Среди всех положительных элементов идеала I выберем наименьший элемент, обозначим его m. Понятно, что $m\mathbb{Z} \subseteq I$. Покажем, что $I = m\mathbb{Z}$.

Пусть $x \in I$. Разложим число x по модулю m : $x = qm + r$, $0 \leqslant r < m$. Тогда $r = x - qm \in I$, из чего следует в силу минимальности числа m, что $r = 0$. Поэтому каждый элемент из идеала I имеет вид zm, $z \in Z$, т.е. $I = m\mathbb{Z}$. $\square$

Следующее предложение доказывается совершенно аналогично, с учетом предложения 2.27.

Предложение 2.37. Если K — поле, то в кольце многочленов $K[x]$ каждый идеал является главным.

Целостное кольцо, в котором каждый идеал является главным, называется *кольцом главных идеалов*. Важными примерами колец главных идеалов служат кольца $\mathbb{Z}$ и $K[x]$.

Предложение 2.38. Диофантово уравнение первой степени:

$$a_1x_1 + \ldots + a_nx_n = b, \quad a_1, \ldots, a_n, b \in \mathbb{Z}, \quad a_1^2 + \ldots + a_n^2 > 0,$$

имеет решение тогда и только тогда, когда $d = (a_1, \ldots, a_n)$ делит b.

Доказательство. Понятно, что если рассматриваемое диофантово уравнение имеет решение, то $d|b$.

Пусть $d|b$. Так как множество:

$$I = \{a_1x_1 + \ldots + a_nx_n \mid x_1, \ldots, x_n \in \mathbb{Z}\}$$

является идеалом в кольце $\mathbb{Z}$, то из предложения 2.36 следует, что этот идеал порождается некоторым элементом $a \in \mathbb{Z}$: $I = a\mathbb{Z}$. При этом $a|a_i$ для любого $i = 1, \ldots, n$. Поэтому $a|d$. С другой стороны, так как $d|a_i$, $i = 1, \ldots, n$, то для любых $x_1, \ldots, x_n \in \mathbb{Z}$ $d|(a_1x_1 + \ldots + a_nx_n)$. Из равенства $I = a\mathbb{Z}$ следует, что $d|a$. Поэтому $a = \pm d$ и исходное диофантово уравнение имеет решение.

2.4.8. Китайская теорема об остатках

Два элемента a и b кольца R называются *сравнимыми по модулю идеала* I (обозначается $a \equiv b \pmod{I}$), если $a - b \in I$.

Теорема 2.5 (**китайская теорема об остатках**)**.** Пусть R — коммутативное кольцо с единицей и $I_1,\ldots,I_n$ — такие идеалы кольца R, что $I_i + I_j = R$ для всех $i \neq j$. Тогда для любого набора элементов $x_1, \ldots, x_n \in R$ найдется такой элемент $x \in R$, что $x \equiv x_i \pmod{I_i}$ для любого $i = 1, \ldots, n$.

Доказательство. Проведем доказательство с помощью индукции по n. Пусть $n = 2$. В силу того, что $I_1 + I_2 = R$ и R — кольцо с единицей, найдутся такие элементы $i_1 \in I_1$, $i_2 \in I_2$, что $i_1 + i_2 = 1$. Тогда в качестве искомого x можно выбрать такой элемент: $x = x_2 i_1 + x_1 i_2$.

Предположим, что теорема верна для любого $k < n$ ($n \geqslant 3$). Докажем для $k = n$. Зафиксируем некоторое s, $1 \leqslant s \leqslant n$. Для каждой из пар идеалов I_s и I_k, $k = 1, \ldots, n$, $k \neq s$, найдется такая пара элементов $a_k \in I_s$, $b_k \in I_k$, что $a_k + b_k = 1$. Понятно, что:

$$\prod_{\substack{1 \leqslant k \leqslant n \\ k \neq s}} (a_k + b_k) = 1.$$

Из определения понятия идеала следует, что это произведение принадлежит идеалу:

$$I_s + \prod_{\substack{1 \leqslant k \leqslant n \\ k \neq s}} I_k.$$

Так как единица принадлежит данному идеалу, то:

$$I_s + \prod_{\substack{1 \leqslant k \leqslant n \\ k \neq s}} I_k = R.$$

В силу справедливости теоремы при $n = 2$, найдется такой элемент $y_s \in R$, что:

$$y_s \equiv 1 \pmod{I_s}, \quad y_s \equiv 0 \pmod{\prod_{\substack{1 \leqslant k \leqslant n \\ k \neq s}} I_k}.$$

Тогда в качестве искомого элемента x можно взять такой:

$$x = \sum_{s=1}^{n} y_s x_s. \qquad \square$$

Рассмотрим еще одну формулировку китайской теоремы об остатках.

Теорема 2.6. Пусть R — коммутативное кольцо с единицей и $I_1,\ldots,I_n$ — такие идеалы кольца R, что $I_i + I_j = R$ для всех $i \neq j$. Тогда отображение:

$$\varphi : R \to R/I_1 \oplus \ldots \oplus R/I_n,$$

определенное правилом $\varphi(x) = (x + I_1, \ldots, x + I_n)$, является сюрьективным гомоморфизмом с ядром $Ker\ \varphi = I_1 \cap \ldots \cap I_n$, т.е.:

$$R/(I_1 \cap \ldots \cap I_n) \cong R/I_1 \oplus \ldots \oplus R/I_n.$$

Доказательство. Очевидно, что φ является гомоморфизмом. Сюрьективность данного гомоморфизма следует из теоремы 2.5: для любых элементов $x_1, \ldots, x_n \in R$ найдется такой элемент $x \in R$, что $x + I_i = x_i + I_i$ для любого $i = 1, \ldots, n$, т.е.:

$$\varphi(x) = (x_1 + I_1, \ldots, x_n + I_n).$$

Нетрудно видеть, что $Ker\ \varphi = I_1 \cap \ldots \cap I_n$. Осталось применить теорему 2.4. □

Следствие 2.4. Пусть натуральное число n имеет каноническое разложение $n = p_1^{\alpha_1} \ldots p_k^{\alpha_k}$. Тогда:

$$\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{\alpha_1}} \oplus \ldots \oplus \mathbb{Z}_{p_k^{\alpha_k}}.$$

Доказательство. В данном случае в роли кольца R выступает кольцо целых чисел $\mathbb{Z}$, а в роли идеалов — $p_1^{\alpha_1}\mathbb{Z},\ldots,p_k^{\alpha_k}\mathbb{Z}$. Гомоморфизм:

$$\varphi : \mathbb{Z} \to \mathbb{Z}_{p_1^{\alpha_1}} \oplus \ldots \oplus \mathbb{Z}_{p_k^{\alpha_k}}$$

определен правилом:

$$\varphi(x) = (x + p_1^{\alpha_1}\mathbb{Z}, \ldots, x + p_k^{\alpha_k}\mathbb{Z}).$$

Так как числа $p_i^{\alpha_i}$ и $p_j^{\alpha_j}$ являются взаимно простыми для любых $i \neq j$, то найдутся такие u и v (зависящие от i и j), что:

$$up_i^{\alpha_i} + vp_j^{\alpha_j} = 1.$$

Поэтому $1 \in p_i^{\alpha_i}\mathbb{Z} + p_j^{\alpha_j}\mathbb{Z}$, что означает равенство $\mathbb{Z} = p_i^{\alpha_i}\mathbb{Z} + p_j^{\alpha_j}\mathbb{Z}$. Таким образом, все условия теоремы 2.6 выполнены. При этом:

$$Ker\ \varphi = p_1^{\alpha_1}\mathbb{Z} \cap \ldots \cap p_k^{\alpha_k}\mathbb{Z} = p_1^{\alpha_1} \ldots p_k^{\alpha_k}\mathbb{Z} = n\mathbb{Z}.$$ □

2.5. Поля

Определение 2.17. *Полем* называется ассоциативное коммутативное кольцо с единицей $1 \neq 0$, в котором каждый ненулевой элемент обратим (по умножению).

Отметим следующие свойства поля.

1. В поле нет делителей нуля, так как если $ab = 0$ и $a \neq 0$, то $a^{-1}(ab) = b = 0$.

2. В поле F множество всех ненулевых элементов образует мультипликативную абелеву группу: $F^* = F \setminus \{0\}$ (данная группа называется *мультипликативной группой поля*). Из этого следует, что в поле существует единственная единица и для любого ненулевого элемента существует единственный обратный по умножению элемент.

3. Любое поле не содержит собственных идеалов, т.е. любой идеал поля F либо нулевой, либо совпадает с самим F. Действительно, пусть I — ненулевой идеал поля F. Тогда найдется ненулевой элемент $a \in I$. Из этого следует, что $1 = a^{-1}a \in I$. Поэтому для любого $x \in F$ следует, что $x = x \cdot 1 \in I$, т.е. $I = F$.

Для любых элементов $a, b \in F$, $b \neq 0$, произведение ab^{-1} обычно записывается в виде дроби $\dfrac{a}{b}$ или a/b. При этом сохраняются обычные правила обращения с дробями:

$$\frac{a}{b} = \frac{c}{d} \Leftrightarrow ad = bc,$$

$$\frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},$$

$$-\frac{a}{b} = \frac{-a}{b} = \frac{a}{-b},$$

$$\left(\frac{a}{b}\right)^{-1} = \frac{b}{a}, \quad a \neq 0,$$

где $a, b, c, d \in F$, причем $b, d \neq 0$. Эти правила легко доказываются. Докажем, например первое правило. Равенство $a/b = c/d$ равносильно равенству $ab^{-1} = cd^{-1}$. Домножив обе части данного равенства на bd, получим $ad = bc$.

2.5.1. Простые идеалы

Определение 2.18. Идеал I коммутативного кольца с единицей R называется *простым*, если фактор-кольцо R/I является целостным, т.е. в данном случае не содержит делителей нуля.

Например, в кольце целых чисел простыми будут главные идеалы, порожденные простыми числами.

Предложение 2.39. Идеал I коммутативного кольца с единицей R является простым тогда и только тогда, когда из того, что $ab \in I$ следует, что либо $a \in I$, либо $b \in I$.

Определение 2.19. Собственный идеал I кольца R называется *максимальным*, если он не содержится ни в каком другом идеале, кроме самого R.

Например, главные идеалы вида $p\mathbb{Z}$, где p — простые числа, являются максимальными в кольце $\mathbb{Z}$.

Предложение 2.40. (i) Идеал I коммутативного кольца с единицей R является максимальным тогда и только тогда, когда фактор-кольцо R/I является полем.

(ii) Всякий максимальный идеал коммутативного кольца с единицей является простым.

Доказательство. (i) Пусть $a \in R \setminus I$. Рассмотрим идеал $aR+I$ кольца R. Так как R — кольцо с единицей, то $a \in aR+I$. Также для произвольного кольца R выполнено строгое включение $I \subset aR + I$. В силу максимальности идеала I будет выполнено равенство $R = aR + I$. Поэтому найдутся такие элементы $b \in R$ и $i \in I$, что $1 = ab + i$, из чего следует, что:

$$1 + I = ab + I = (a + I)(b + I).$$

Следовательно, все ненулевые элементы фактор-кольца R/I обратимы, поэтому данное фактор-кольцо является полем.

Обратно, пусть для некоторого собственного идеала I фактор-кольцо R/I является полем. Предположим, что I содержится в некотором собственном идеале J кольца R и $I \neq J$.

Очевидно, что $1 \notin J$. Пусть $a \in J \setminus I$. В силу того, что J является идеалом, выполнено включение $JR \subseteq J$. Поэтому для любого $b \in R$:

$$(a+I)(b+I) \neq 1+I.$$

Таким образом, ненулевой элемент $a+I$ поля R/I не имеет обратного. Противоречие.

(ii) Так как в поле нет делителей нуля, то I является простым идеалом. □

2.5.2. Подполе. Поле частных

Определение 2.20. Подмножество K поля F называется *подполем поля* F, если оно само является полем относительно операций, определенных в F. При этом говорят, что поле F является *расширением поля* K.

Например, поле $\mathbb{Q}$ является подполем в $\mathbb{R}$.

Поскольку для любого поля F множество F с операцией сложения $(F,+)$ является аддитивной группой с 0, а $(F \setminus \{0\},\cdot)$ — мультипликативной группой с 1, то ноль и единица любого подполя будут совпадать с 0 и 1 поля F.

Предложение 2.41 (критерий подполя). Подмножество K поля F, содержащее хотя бы один ненулевой элемент, является подполем тогда и только тогда, когда выполнены следующие условия:

(i) для любых $x, y \in K$ выполнено $x-y \in K$, $xy \in K$;

(ii) для любого $x \in K \setminus \{0\}$ выполнено $x^{-1} \in K$.

Доказательство следует из предложения 2.28 и определения поля. □

Следствие 2.5. 1. Отношение «быть подполем» транзитивно на любом множестве полей.

2. Пересечение любого семейства подполей поля F является его подполем.

Предложение 2.42 (критерий конечного подполя). Конечное подмножество K поля F, содержащее хотя бы один ненулевой элемент, является подполем тогда и только тогда, когда выполнено условие:

для любых $x, y \in K$ выполнено $x + y \in K$, $xy \in K$.

Доказательство следует из предложения 2.16, так как в этом случае $(K, +)$ будет являться аддитивной абелевой группой, а $(K \setminus \{0\}, \cdot)$ — мультипликативной абелевой группой. $\square$

Определение 2.21. Поле F называют полем частных кольца R, если:

а) существует изоморфное вложение $\varphi : R \to F$;

б) каждый элемент поля F имеет вид $\varphi(a)\varphi(b)^{-1}$, $a \in R$, $b \in R \setminus \{0\}$.

Например, поле $\mathbb{Q}$ является полем частных кольца $\mathbb{Z}$.

Из пункта а) определения 2.21 следует, что если ненулевое кольцо R имеет поле частных, то данное кольцо является коммутативным и не имеет делителей нуля. В следующей теореме показано, что данное условие является не только необходимым, но и достаточным.

Теорема 2.7. Если R — ненулевое коммутативное кольцо без делителей нуля, то для него существует поле частных.

Доказательство. На множестве $M = R \times (R \setminus \{0\})$ определим бинарное отношение следующим образом:

$$(a, b) \sim (c, d) \Leftrightarrow ad = bc.$$

Очевидно, что данное бинарное отношение является рефлексивным и симметричным. Покажем свойство транзитивности данного отношения. Пусть $(a, b) \sim (c, d)$, $(c, d) \sim (e, f)$. Умножим равенство $ad = bc$ на f, а равенство $cf = de$ — на b : $adf = bcf$, $bcf = bde$. Поэтому выполнено равенство $adf = bde$, из которого следует, что $d(af - be) = 0$. Так как $d \neq 0$ и в кольце R нет делителей нуля, то $af = be$. Поэтому $(a, b) \sim (e, f)$ и

отношение $\sim$ является отношением эквивалентности на множестве M, которое индуцирует на данном множестве разбиение на непересекающиеся классы.

Обозначим через $Q = Q(R)$ множество всех классов эквивалентности или, что то же самое, Q есть фактор-множество $M/\sim$ множества M по отношение эквивалентности $\sim$. Будем обозначать символом $[a, b]$ класс, порожденный элементом (a, b). По определение $[a, b] = [c, d] \Leftrightarrow ad = bc$.

Определим на множестве Q операции сложения и умножения следующим образом:

$$[a, b] + [c, d] = [ad + bc, bd], \quad [a, b] \cdot [c, d] = [ac, bd].$$

Покажем корректность данных операций. Так как в кольце R нет делителей нуля, то $bd \neq 0$. Пусть $(a, b) \sim (x, y)$ и $(c, d) \sim (s, t)$. Нужно показать, что:

$$[x, y] + [s, t] = [xt + ys, yt] = [ad + bc, bd]$$

$$(\Leftrightarrow (xt + ys)bd = yt(ad + bc)),$$

$$[x, y] \cdot [s, t] = [xs, yt] = [ac, bd] \quad (\Leftrightarrow xsbd = ytac).$$

Так как $ay = bx$, $ct = ds$, то:

$$(xt + ys)bd = bxtd + dsyb = aytd + ctyb = yt(ad + bc),$$

$$xsbd = aysd = ayct.$$

Теперь покажем, что $(Q, +, \cdot)$ — поле. Из соотношений:

$$([a, b]+[c, d])+[e, f] = [adf+bcf+bde, bdf] = [a, b]+([c, d]+[e, f])$$

следует ассоциативность операции сложения. Ассоциативность умножения очевидна. Коммутативность операций сложения и умножения также очевидна. Так как для любых $s, t \in R \setminus \{0\}$ выполнены равенства $[0, s] = [0, t]$, $[s, s] = [t, t]$, то:

$$[a, b] + [0, s] = [as, bs] = [a, b], \quad [a, b] \cdot [s, s] = [as, bs] = [a, b].$$

Поэтому $[0, s]$ — нулевой элемент относительно сложения, $[s, s]$ — единичный элемент относительно умножения. Далее:

$$-[a, b] = [-a, b], \quad [a, b]^{-1} = [b, a]$$

при условии, что $a \neq 0$. Дистрибутивность следует из равенств:

$$([a,b]+[c,d])\cdot[e,f]=[ade+bce,bdf],$$

$$[a,b]\cdot[e,f]+[c,d]\cdot[e,f]=[aedf+bfce,bfdf]=[(ade+bce)f,(bdf)f].$$

Таким образом, $(Q,+,\cdot)$ — поле. Определим отображение $\varphi : R \to Q$, определив для фиксированного $s \in R \setminus \{0\}$:

$$\varphi(x)=[xs,s], \quad x \in R.$$

Нетрудно проверить, что данное отображение является гомоморфизмом колец, причем $Ker\ \varphi = \{0\}$. Поэтому φ — изоморфное вложение.

Так как:

$$[a,b]=[as,bs]=[as,s]\cdot[s,bs]=[as,s]\cdot[bs,s]^{-1}=\varphi(a)\varphi(b)^{-1},$$

то Q — поле частных кольца R. □

Пример 2.21. 1. Пусть F — поле, R — его ненулевое подкольцо. Рассмотрим множество $K=\{ab^{-1} \mid a \in R,\ b \in R \setminus \{0\}\}$. Так как

$$ab^{-1}+cd^{-1}=ad(bd)^{-1}+bc(bd)^{-1}=(ad+bc)(bd)^{-1},$$

то из предложения 2.41 следует, что K — подполе в F, являющееся полем частных кольца R (определение 2.21). При этом нетрудно проверить, что K — пересечение всех подполей в F, содержащих R.

2. Кольцо многочленов $F[x]$ над полем F является коммутативным кольцом без делителей нуля. По теореме 2.7 для него существует поле частных:

$$F(x)=\left\{\frac{f(x)}{g(x)} \mid f(x),g(x)\in F[x],\ g(x)\neq 0\right\},$$

которое называется полем рациональных дробей. Каждая такая рациональная дробь записывается (притом многими способами) в виде $\dfrac{f(x)}{g(x)}$. По определению $\dfrac{f(x)}{g(x)}=\dfrac{f_1(x)}{g_1(x)}$ тогда и только тогда, когда $f(x)g_1(x)=f_1(x)g(x)$.

Теорема 2.8. Пусть $\alpha : R_1 \to R_2$ — изоморфизм ненулевых коммутативных колец без делителей нуля, F_i — поле частных кольца R_i, $\varphi_i : R_i \to F_i$ — изоморфное вложение, удовлетворяющее определению 2.21, $i = 1, 2$. Тогда существует такой изоморфизм $\beta : F_1 \to F_2$, что $\beta(\varphi_1(a)) = \varphi_2(\alpha(a))$ для любого $a \in R_1$, т.е. коммутативна диаграмма:

$$\begin{array}{ccc} R_1 & \xrightarrow{\alpha} & R_2 \\ {\scriptstyle\varphi_1}\downarrow & & \downarrow{\scriptstyle\varphi_2} \\ F_1 & \xrightarrow[\beta]{} & F_2 \end{array}$$

Доказательство. Зададим отображение $\beta : F_1 \to F_2$, положив:

$$\beta(\varphi_1(a)\varphi_1(b)^{-1}) = \varphi_2(\alpha(a))\varphi_2(\alpha(b))^{-1}, \quad a \in R_1,\ b \in R_1 \setminus \{0\}.$$

Покажем корректность данного отображения. Пусть:

$$\varphi_1(a)\varphi_1(b)^{-1} = \varphi_1(\widetilde{a})\varphi_1(\widetilde{b})^{-1}.$$

Тогда $\varphi_1(a)\varphi_1(\widetilde{b}) = \varphi_1(\widetilde{a})\varphi_1(b)$, $\varphi_1(a\widetilde{b} - \widetilde{a}b) = 0$, $a\widetilde{b} = \widetilde{a}b$. Поэтому $\alpha(a\widetilde{b}) = \alpha(\widetilde{a}b)$,

$$\varphi_2(\alpha(a))\varphi_2(\alpha(\widetilde{b})) = \varphi_2(\alpha(a)\alpha(\widetilde{b})) = \varphi_2(\alpha(a\widetilde{b})) =$$
$$= \varphi_2(\alpha(\widetilde{a}b)) = \varphi_2(\alpha(\widetilde{a})\alpha(b)) = \varphi_2(\alpha(\widetilde{a}))\varphi_2(\alpha(b)),$$
$$\varphi_2(\alpha(a))\varphi_2(\alpha(b))^{-1} = \varphi_2(\alpha(\widetilde{a}))\varphi_2(\alpha(\widetilde{b}))^{-1}.$$

Значит $\beta(\varphi_1(a)\varphi_1(b)^{-1}) = \beta(\varphi_1(\widetilde{a})\varphi_1(\widetilde{b})^{-1})$.

Нетрудно проверить, что β — изоморфизм полей. При этом $\varphi_1(a) = \varphi_1(ab)\varphi_1(b)^{-1}$. Поэтому:

$$\beta(\varphi_1(a)) = \beta(\varphi_1(ab)\varphi_1(b)^{-1}) = \varphi_2(\alpha(a)). \qquad \square$$

Следствие 2.6. Пусть F_i — поле частных ненулевого коммутативного кольца R без делителей нуля, $\varphi_i : R \to F_i$ — изоморфное вложение, удовлетворяющее определению 2.21, $i = 1, 2$. Тогда существует такой изоморфизм $\beta : F_1 \to F_2$, что $\beta(\varphi_1(a)) = \varphi_2(a)$ для любого $a \in R$.

Следствие 2.7. Если в условии следствия 2.6 φ_1 и φ_2 — тождественные вложения, то поля F_1 и F_2 изоморфны над R, т.е. существует такой изоморфизм $\beta : F_1 \to F_2$, что $\beta(a) = a$ для любого $a \in R$.

2.5.3. Простые поля. Характеристика поля

Определение 2.22. Поле F называется *простым*, если оно не содержит собственных подполей, кроме самого F.

Пример 2.22. 1. Кольцо вычетов $\mathbb{Z}_p$ по простому модулю p является простым подполем, так как $(\mathbb{Z}_p, +)$ является циклической группой простого порядка p и по теореме Лагранжа не содержит собственных подгрупп.

2. Поле $\mathbb{Q}$ является простым.

Теорема 2.9. Каждое поле содержит одно и только одно простое подполе. Это подполе изоморфно либо $\mathbb{Q}$, либо $\mathbb{Z}_p$ для некоторого простого p.

Доказательство. Пусть F_0 — пересечение всех подполей в F. Поле F_0 не имеет собственных подполей, т.е. поле F_0 является простым. При этом поле F не содержит никаких других простых подполей. Действительно, если K — некоторое простое подполе в F, отличное от F_0, то пересечение $K \cap F_0$ является собственным подполем в K и F_0. Противоречие. Поэтому простое подполе единственно.

Покажем вторую часть утверждения. Так как F_0 — поле, то $1 \in F_0$ и:

$$n \cdot 1 = \underbrace{1 + \ldots + 1}_{n} \in F_0, \quad (-n) \cdot 1 = \underbrace{-1 - \ldots - 1}_{n} \in F_0,$$

где n — произвольное натуральное число. Поэтому:

$$(s \cdot 1) + (t \cdot 1) = (s + t) \cdot 1, \quad (s \cdot 1) \cdot (t \cdot 1) = (st) \cdot 1, \quad s, t \in \mathbb{Z}.$$

Таким образом, циклическая абелева группа $\langle 1 \rangle$, образованная единицей, с добавлением операции умножения является коммутативным кольцом с единицей $(\langle 1 \rangle, +, \cdot)$, содержащееся в поле F_0. Построим кольцевой гомоморфизм $\varphi : \mathbb{Z} \to F_0$ следующим образом: $\varphi(n) = n \cdot 1$, $n \in \mathbb{Z}$. Из предложений 2.34 и 2.36 следует, что ядро данного гомоморфизма имеет вид $Ker\ \varphi = m\mathbb{Z}$ для некоторого неотрицательного целого числа m. Относительно числа m возможны следующие случаи.

$m = 0$. Тогда $Ker\ \varphi = \{0\}$, что означает изоморфное вложение кольца целых чисел $\mathbb{Z}$ в поле F_0. При этом для любого ненулевого числа $t \in \mathbb{Z}$ для элемента $t \cdot 1 \in F_0$ существует обратный в F_0 элемент $(t \cdot 1)^{-1}$. Поэтому для любых $s, t \in \mathbb{Z}$, $t \neq 0$, элемент $(s \cdot 1)/(t \cdot 1) \in F_0$. Пусть:

$$K = \{(s \cdot 1)(t \cdot 1)^{-1} \mid s \in \mathbb{Z},\ t \in \mathbb{Z} \setminus \{0\}\} \subseteq F_0.$$

Из примера 2.21 следует, что K — поле частных кольца $\varphi(\mathbb{Z})$. Так как F_0 — простое поле, то $K = F_0$. Из теоремы 2.8 следует, что поля $\mathbb{Q}$ и F_0 изоморфны:

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\varphi} & \varphi(\mathbb{Z}) \\ {\scriptstyle\varphi_1}\downarrow & & \downarrow{\scriptstyle\varphi_2} \\ \mathbb{Q} & \xrightarrow[\beta]{} & F_0 \end{array}$$

$m > 0$. В этом случае из теоремы 2.4 следует, что кольцо вычетов $\mathbb{Z}_m$ изоморфно вкладывается в поле F_0. Поскольку в любом поле нет делителей нуля, то это вложение возможно только в случае, если $m = p$ — простое число. При этом кольцо $\mathbb{Z}_p$ является полем. Осталось вспомнить, что поле F_0 является простым, поэтому изоморфный образ поля $\mathbb{Z}_p$ совпадает с F_0. □

Если простое подполе в F изоморфно $\mathbb{Q}$, то говорят, что поле F имеет *нулевую характеристику* (обозначается $char\ F = 0$); если же его простое подполе изоморфно $\mathbb{Z}_p$, то говорят, что поле F имеет *положительную характеристику, равную* p (обозначается $char\ F = p$).

2.5.4. Расширение полей

Если K — подполе поля F, то говорят, что поле F является *расширением поля* K. Так как в этом случае можно рассматривать поле F как векторное пространство над полем K, то размерность этого векторного пространства обозначим через $[F : K]$ и будем называть *степенью* расширения. Размерность может быть и бесконечной.

Теорема 2.10 (**о башне полей**)**.** В башне расширений:

$$K \subset F \subset E$$

степень $[E : K]$ конечна тогда и только тогда, когда конечны степени $[F : K]$ и $[E : F]$. При выполнении последнего условия выполнено равенство:

$$[E : K] = [F : K][E : F].$$

Доказательство. Пусть степени $[F : K]$ и $[E : F]$ конечны. Обозначим через $e_1, \ldots, e_m$ — базис пространства E над F, а через $f_1, \ldots, f_n$ — базис пространства F над K. Пусть $x \in E$. Тогда найдутся такие $\alpha_1,\ldots,\alpha_m \in F$, что $x = \sum_{i=1}^{m} \alpha_i e_i$. В свою очередь для любого $i = 1, \ldots, m$ найдутся такие $\beta_{i1},\ldots,\beta_{in} \in K$, что $\alpha_i = \sum_{j=1}^{n} \beta_{ij} f_j$. Следовательно, $x = \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{ij} e_i f_j$. Поэтому линейная оболочка элементов $e_i f_j$ порождает векторное пространство E над K. Покажем линейную независимость элементов $e_i f_j$. Пусть:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{ij} e_i f_j = 0, \quad \beta_{ij} \in K.$$

Тогда:

$$0 = \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{ij} e_i f_j = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \beta_{ij} f_j \right) e_i.$$

Так как $e_1,\ldots,e_m$ — базис пространства E над F, то для любого $i = 1, \ldots, m$:

$$\sum_{j=1}^{n} \beta_{ij} f_j = 0.$$

Поэтому все $\beta_{ij} = 0$ в силу того, что $f_1, \ldots, f_n$ — базис пространства F над K. Таким образом, $[E : K] = [F : K][E : F]$.

Обратное утверждение очевидно. □

Пусть K — поле, F — его расширение и $\alpha \in F$. Под обозначением $K(\alpha)$ будем понимать наименьшее подполе в F, содержащее K и α. При этом говорят, что поле $K(\alpha)$ получено

путем *присоединения к K элемента α* и называется *простым расширением* поля K. Поле $K(\alpha)$ является пересечением всех подполей в F, содержащих K и α. Нетрудно видеть, что поле $K(\alpha)$ будет состоять из всех элементов вида $f(\alpha)/g(\alpha)$, где $f(x), g(x) \in K[x]$ и $g(\alpha) \neq 0$.

Пусть $\alpha_1, \ldots, \alpha_n \in F$. Аналогично будем обозначать через $K(\alpha_1, \ldots, \alpha_n)$ наименьшее подполе в поле F, содержащее K и $\alpha_1, \ldots, \alpha_n$. Данное поле будет состоять из всех дробей вида:

$$\frac{f(\alpha_1, \ldots, \alpha_n)}{g(\alpha_1, \ldots, \alpha_n)},$$

где $f, g \in K[x_1, \ldots, x_n]$ и $g(\alpha_1, \ldots, \alpha_n) \neq 0$.

Пример 2.23. Рассмотрим поле $\mathbb{Q}$ и его расширение — поле $\mathbb{R}$. Рассмотрим поле $\mathbb{Q}(\sqrt{p})$, p — простое.

Заметим, что число $\sqrt{p}$ иррациональное. Действительно, предположим, что для некоторых $m \in \mathbb{Z}$, $n \in \mathbb{N}$ выполнено равенство $\sqrt{p} = \frac{m}{n}$. Тогда $m^2 = pn^2$ и каноническое разложение числа m^2 содержит число p в четной степени, а pn^2 — в нечетной. Противоречие.

Так как для любого натурального n имеет место равенство $(\sqrt{p})^n = a + b\sqrt{p}$, где $a, b \in \mathbb{Z}$, то поле $\mathbb{Q}(\sqrt{p})$ будет состоять из всех элементов вида $(a + b\sqrt{p})/(c + d\sqrt{p})$, где $a, b, c, d \in \mathbb{Q}$, $c^2 + d^2 \neq 0$. Домножив данную дробь на сопряженную дробь $(c - d\sqrt{p})/(c - d\sqrt{p})$, получим элемент вида $(x + y\sqrt{p})$, $x, y \in \mathbb{Q}$. Поэтому:

$$\mathbb{Q}(\sqrt{p}) = \{a + b\sqrt{p} \mid a, b \in \mathbb{Q}\}.$$

Определение 2.23. Пусть F — подполе поля E. Элемент $\alpha \in E$ называется *алгебраическим* над F, если в F существуют элементы $\beta_0, \ldots, \beta_n$, не все равные нулю, что:

$$\beta_0 + \beta_1\alpha + \ldots + \beta_n\alpha^n = 0,$$

иными словами, α является корнем некоторого ненулевого многочлена из $F[x]$. В противном случае элемент α называют *трансцендентным* над F. Расширение E поля F называется *алгебраическим*, если всякий элемент из E алгебраичен над F.

Пример 2.24. 1. Кольцо многочленов $\mathbb{Q}[x]$ счетно. При этом любой ненулевой многочлен из $\mathbb{Q}[x]$ может иметь в поле $\mathbb{R}$ конечное число корней. Поэтому в $\mathbb{R}$ имеется не более чем счетное множество элементов, алгебраических над $\mathbb{Q}$. Поскольку множество $\mathbb{R}$ более чем счетно, то в $\mathbb{R}$ существуют трансцендентные элементы, например, число π и основание натурального логарифма e.

2. Поле рациональных функций $F(x)$ является расширением поля F. Любой элемент $f \in F(x) \setminus F$ является трансцендентным над F.

Предложение 2.43. Если $F \subset E$ и $\alpha \in E$ — алгебраический элемент над F, то в $F[x]$ существует единственный унитарный неприводимый над F многочлен $p(x)$, корнем которого является α. При этом если для некоторого $f(x) \in F[x]$ выполнено равенство $f(\alpha) = 0$, то $p(x) \mid f(x)$.

Доказательство. Рассмотрим множество:

$$I = \{f(x) \in F[x] \mid f(\alpha) = 0\}.$$

Нетрудно проверить, что I является идеалом кольца $F[x]$. По предложению 2.37 найдется такой многочлен $p(x) \in F[x]$, что $I = (p(x))$, т.е. I порождается многочленом $p(x)$. При этом $f(x) \in I$ тогда и только тогда, когда $p(x) \mid f(x)$.

Пусть теперь $p(x) = u(x)v(x)$ для некоторых многочленов $u(x), v(x) \in F[x]$, степень которых не меньше единицы и строго меньше степени многочлена $p(x)$. Так как $0 = p(\alpha) = u(\alpha)v(\alpha)$, но $u(\alpha) \neq 0$, $v(\alpha) \neq 0$, то $u(\alpha)$ и $v(\alpha)$ являются делителями нуля в поле E. Противоречие. Поэтому многочлен $p(x)$ неприводим над $F[x]$. □

Определение 2.24. Многочлен $p(x)$ из предложения 2.43 называется *минимальным многочленом* элемента α над F и обозначается $Irr(\alpha, F, x)$.

Предложение 2.44. Всякое конечное расширение E поля F алгебраично над F.

Доказательство. Пусть $n = [E : F]$. Тогда $n+1$ элементов вида $1, \alpha, \ldots, \alpha^n$ линейно зависимы над F. Таким образом, $\sum_{i=0}^{n} \beta_i \alpha^i = 0$ для некоторых $\beta_i \in F$ и элемент α является корнем многочлена $\sum_{i=0}^{n} \beta_i x^i \in F[x]$. □

Теорема 2.11 (классификация простых расширений поля). Пусть $F \subset E$ и $\alpha \in E$.

1. Если α — алгебраический элемент над F, то:

$$F(\alpha) = F[\alpha] \cong F[x]/(Irr(\alpha, F, x)).$$

Более того, поле $F(\alpha)$ конечно над F, причем степень $[F(\alpha) : F]$ равна степени многочлена $Irr(\alpha, F, x)$:

$$F(\alpha) = \langle 1,\ \alpha,\ \alpha^2, \ldots,\ \alpha^{n-1} \rangle_F,$$

где $n = \deg\ Irr(\alpha, F, x)$.

2. Если α — трансцендентный элемент над F, то $F(\alpha) \cong F(x)$.

Доказательство. 1. Сначала покажем, что выполнено равенство $F(\alpha) = F[\alpha]$. Пусть $f(x) \in F[x]$ и $f(\alpha) \neq 0$. Тогда многочлены $f(x)$ и $Irr(\alpha, F, x)$ взаимно просты и найдутся такие многочлены $u(x), v(x) \in F[x]$, что:

$$u(x)f(x) + v(x)Irr(\alpha, F, x) = 1.$$

Так как $u(\alpha)f(\alpha) = 1$, то элемент $f(\alpha)$ обратим в $F[\alpha]$, поэтому кольцо $F[\alpha]$ является полем, из чего следует, что $F(\alpha) = F[\alpha]$.

Далее, определим отображение $\varphi : F[x] \to F[\alpha]$, положив $\varphi(f(x)) = f(\alpha)$. Так как φ — гомоморфизм колец, то по теореме 2.4 справедлив изоморфизм $F[x]/Ker\ \varphi \cong Im\ \varphi$, причем $Im\ \varphi = F[\alpha]$, а из предложения 2.43 следует, что выполнено $Ker\ \varphi = (Irr(\alpha, F, x))$.

Заметим, что отображение $\psi : F[x]/(Irr) \to F[\alpha]$, где $Irr = Irr(\alpha, F, x)$, определенное правилом $\psi(f(x) + (Irr)) = \varphi(f(x))$, является изоморфизмом, причем $\psi(x + (Irr)) = \alpha$.

Покажем последнее утверждение данного пункта. Пусть $n = \deg\ Irr(\alpha, F, x)$. Понятно, что элементы

$$1,\ \alpha,\ \alpha^2, \ldots,\ \alpha^{n-1}$$

линейно независимы в пространстве $F[\alpha]$ над F. Покажем, что $F[\alpha]$ является линейной оболочкой данных элементов. Пусть $f(x) \in F[x]$. Разложим $f(x)$ по модулю многочлена $Irr(\alpha, F, x)$:

$$f(x) = q(x)Irr(\alpha, F, x)+r(x), \quad q(x), r(x) \in F[x], \quad \deg\, r(x) < n.$$

Осталось заметить, что $f(\alpha) = r(\alpha)$.

2. В этом случае ядро гомоморфизма $\varphi : F[x] \to F[\alpha]$, определенного в предыдущем пункте, нулевое: $Ker\ \varphi = \{0\}$. Поэтому по теореме 2.4 $F[x] \cong F[\alpha]$. В силу примера 2.21 поле частных K кольца $F[\alpha]$ — это пересечение всех подполей поля $F(\alpha)$, содержащих $F[\alpha]$. В силу определения поля $F(\alpha)$, $F(\alpha) = K$. При этом полем частных кольца $F[x]$ является поле рациональных функций $F(x)$. По теореме 2.8 $F(\alpha) \cong F(x)$. □

Следствие 2.8. Для любого расширения полей $F \subset E$ все элементы поля E, алгебраические над F, образуют подполе в E.

Доказательство. Пусть α, β — алгебраические элементы поля E над F. Из теоремы 2.11 следует, что степени расширений $[F(\alpha) : F]$ и $[F(\alpha, \beta) : F(\alpha)]$ конечны, поэтому конечна степень расширения $[F(\alpha, \beta) : F]$ (теорема 2.10). Из предложения 2.44 следует, что расширение $F \subset F(\alpha, \beta)$ алгебраично, поэтому $\alpha - \beta$, $\alpha\beta$, α^{-1} — алгебраические элементы поля E над F. Осталось применить критерий подполя (предложение 2.41). □

Следствие 2.9. В башне расширений $K \subset F \subset E$ каждый этаж алгебраичен тогда и только тогда, когда E алгебраично над K.

Доказательство. Если E алгебраично над K, то очевидно, что E алгебраично над F и F алгебраично над K.

Обратно. Пусть E алгебраично над F и F алгебраично над K. Пусть $\alpha \in E$. Тогда α является корнем ненулевого многочлена $f_0 + f_1x + \ldots f_nx^n$, $f_i \in F$, $i = 0, \ldots, n$. По теоремам 2.10 и 2.11 расширения:

$$K \subset K(f_0, \ldots, f_n), \quad K(f_0, \ldots, f_n) \subset K(f_0, \ldots, f_n, \alpha)$$

конечны, поэтому конечно расширение $K \subset K(f_0, \ldots, f_n, \alpha)$. Следовательно, из предложения 2.44 следует, что α алгебраичен над K. $\square$

Пример 2.25. В поле $\mathbb{C}$ расширение поля $\mathbb{R}$, порожденное элементом $i \in \mathbb{C}$, совпадает с $\mathbb{C}$. Действительно, $Irr(i, \mathbb{R}, x) = x^2 + 1$. Поэтому:

$$\mathbb{R}(i) = \langle 1, i \rangle_{\mathbb{R}} = \{x + yi \mid x, y \in \mathbb{R}\} = \mathbb{C}.$$

2.5.5. Поля разложения многочлена

Определение 2.25. Поле E называется *полем разложения* многочлена $f(x) \in F[x]$, если $F \subset E$ и над полем E многочлен $f(x)$ раскладывается на линейные множители.

Теорема 2.12. Для любого поля F и любого многочлена $f(x) \in F[x]$, $\deg\ f(x) \geqslant 1$, существует расширение E поля F, в котором $f(x)$ имеет корень.

Доказательство. I шаг. Если найдется такой многочлен $p(x) \in F[x]$, что $p(x) \mid f(x)$, $\deg\ p(x) \geqslant 1$, и $p(x)$ неприводим, то понятно, что корень многочлена $p(x)$ является корнем многочлена $f(x)$. Если же $f(x)$ неприводим над F, то обозначим $p(x) = f(x)$.

Рассмотрим канонический гомоморфизм:

$$\varphi : F[x] \to F[x]/(p(x)) = P.$$

Так как $p(x)$ неприводим над F, то кольцо P является полем (предложение 2.40).

Покажем, что в поле P содержится подполе, изоморфное полю F, а именно, $\varphi(F)$. Для этого достаточно показать, что сужение отображения φ на F является инъективным отображением. Пусть $a, b \in F$, $a \neq b$. Если $\varphi(a) = \varphi(b)$, то $a - b$ принадлежит ядру $Ker\ \varphi = (p(x))$ гомоморфизма φ, порожденного многочленом $p(x)$. Поэтому $p(x) \mid (a - b)$, что невозможно, так как многочлен $a - b$ имеет нулевую степень.

Теперь покажем, что многочлен $\varphi(p) \in \varphi(F)[x]$ имеет корень в поле P. Корнем этого многочлена является элемент кольца

вычетов $\overline{x} = x + (p(x))$:

$$(\varphi(p))(\overline{x}) = \overline{p(x)} = \overline{0}.$$

II шаг. Пусть S — некоторое множество той же мощности, что и $P \setminus \varphi(F)$, причем S не пересекается с F. Рассмотрим множество $E = F \cup S$. Очевидно, что можно продолжить отображение $\varphi : F \to P$ до биекции E в P. Определим на E структуру поля следующим образом. Пусть $x, y \in E$, положим:

$$x + y = \varphi^{-1}(\varphi(x) + \varphi(y)),$$

$$xy = \varphi^{-1}(\varphi(x)\varphi(y)).$$

Так как φ — гомоморфизм отображения F в P, то введенные бинарные операции сложения и умножения совпадают с соответствующими операциями исходного поля F. При этом понятно, что F — подполе поля E и φ — изоморфизм полей P и E.

Пусть $\alpha = \varphi^{-1}(\overline{x})$, $f(x) = \sum_{i=0}^{n} f_i x^i$. Так как для любого $a \in E$ выполнено равенство $a^n = \varphi^{-1}\Big(\big(\varphi(a)\big)^n\Big)$, то:

$$\sum_{i=0}^{n} f_i \alpha^i = \sum_{i=0}^{n} f_i \varphi^{-1}(\overline{x}^i) = \sum_{i=0}^{n} \varphi^{-1}(\varphi(f_i))\varphi^{-1}(\overline{x}^i) =$$

$$= \varphi^{-1}\left(\sum_{i=0}^{n} \varphi(f_i)\overline{x}^i\right) = \varphi^{-1}(\overline{0}) = 0.$$

Поэтому α — корень многочлена $f(x)$. □

Теорема 2.13. Для любого поля F и любого многочлена $f(x) \in F[x]$, $\deg\ f(x) \geqslant 1$, существует поле разложения многочлена $f(x)$.

Доказательство. Пусть $f(x) = \sum_{i=0}^{n} a_i x^i$ и:

$$f(x) = a_n p_1(x)^{k_1} \dots p_s(x)^{k_s} \tag{2.6}$$

— каноническое разложение многочлена $f(x)$ над полем F. Обозначим:

$$d_F(f) = \sum_{\substack{1 \leqslant i \leqslant s \\ \deg\ p_i(x) > 1}} \deg\ p_i(x).$$

Проведем индукцию по $d_F(f)$. Если $d_F(f) = 0$, то многочлен $f(x)$ раскладывается над полем F на линейные множители. Тогда поле F является полем разложения многочлена $f(x)$.

Предположим, что теорема верна для любого многочлена $f(x) \in F[x]$ с условием $d_F(f) < k$. Покажем справедливость теоремы для случая $d_F(f) = k$. Предположим, что в разложении (2.6) $\deg\ p_1(x) > 1$. По теореме 2.12 существует расширение E поля F, в котором многочлен $p_1(x)$ имеет корень. Разложим многочлен $f(x)$ на неприводимые множители над полем E. Тогда $d_E(f) < k$, так как над полем E многочлен $p_1(x)$ имеет корень. Далее применяем предположение индукции. □

Предложение 2.45. Пусть $\sigma : F \to E$ — изоморфизм полей. Продолжим его до отображения колец $\sigma : F[x] \to E[x]$ следующим образом:

$$\sigma\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=0}^{n} \sigma(a_i) x^i, \quad a_i \in F.$$

Тогда:

(i) σ — изоморфизм колец;

(ii) если $f(x), g(x) \in F[x]$, то $f(x) \mid g(x)$ тогда и только тогда, когда $\sigma\big(f(x)\big) \mid \sigma\big(g(x)\big)$; многочлен $f(x)$ неприводим над F тогда и только тогда, когда $\sigma\big(f(x)\big)$ неприводим над E;

(iii) для любого $f(x) \in F[x]$ имеет место изоморфизм колец:

$$F[x]/\big(f(x)\big) \cong E[x]/\Big(\sigma\big(f(x)\big)\Big).$$

Доказательство. (i) Пусть $f(x), g(x) \in F[x]$. Тогда, учитывая (2.3) и (2.4), очевидно, что:

$$\sigma\big(f(x) + g(x)\big) = \sigma\big(f(x)\big) + \sigma\big(g(x)\big),$$
$$\sigma\big(f(x)g(x)\big) = \sigma\big(f(x)\big)\sigma\big(g(x)\big).$$

(ii) следует из (i).

(iii) Пусть $\tau : E[x] \to E[x]/\Big(\sigma\big(f(x)\big)\Big)$ — канонический гомоморфизм. Тогда композиция $\tau \circ \sigma$:

$$F[x] \xrightarrow{\sigma} E[x] \xrightarrow{\tau} E[x]/\Big(\sigma\big(f(x)\big)\Big)$$

также является гомоморфизмом. При этом, учитывая пункт (ii), имеем:

$$Ker\ \tau \circ \sigma = \{g(x) \in F[x]\ :\ f(x) \mid g(x)\} = \big(f(x)\big).$$

Осталось применить теорему 2.4. □

Предложение 2.46. Пусть $F(\alpha)$ — расширение поля F, порожденное корнем α многочлена $Irr(\alpha, F, x)$, $E(\beta)$ — расширение поля E, порожденное корнем β многочлена $Irr(\beta, E, x)$. Если существует такой изоморфизм $\sigma : F \to E$, что:

$$\sigma(Irr(\alpha, F, x)) = Irr(\beta, E, x),$$

то существует такой изоморфизм $\tau : F(\alpha) \to E(\beta)$, что $\tau|_F = \sigma$ и $\tau(\alpha) = \beta$.

Доказательство. По теореме 2.11 существуют изоморфизмы:

$$\tau_1 : F(\alpha) \to F[x]/\big(Irr(\alpha, F, x)\big),$$
$$\tau_2 : E(\beta) \to E[x]/\big(Irr(\beta, E, x)\big),$$

при которых:

$$\tau_1(\alpha) = \overline{x} = x + (Irr(\alpha, F, x)),$$

$$\tau_2(\beta) = \overline{x} = x + (Irr(\beta, E, x)).$$

Так как $\sigma(Irr(\alpha, F, x)) = Irr(\beta, E, x)$, то по предложению 2.45 существует изоморфизм:

$$\varphi : F[x]/\big(Irr(\alpha, F, x)\big) \to E[x]/\big(Irr(\beta, E, x)\big),$$

при котором:

$$\varphi\Big(f(x) + \big(Irr(\alpha, F, x)\big)\Big) = \sigma\big(f(x)\big) + \big(Irr(\beta, E, x)\big).$$

Пусть $\tau = \tau_2^{-1} \circ \varphi \circ \tau_1$. Тогда:

$$\tau(\alpha) = (\tau_2^{-1} \circ \varphi \circ \tau_1)(\alpha) = (\tau_2^{-1} \circ \varphi)\Big(x + \big(Irr(\alpha, F, x)\big)\Big) =$$

$$= \tau_2^{-1}\Big(x + \big(Irr(\beta, E, x)\big)\Big) = \beta.$$

При этом понятно, что $\tau|_F = \sigma$. □

Определение 2.26. Поле разложения E многочлена $f(x) \in F[x]$ называется *минимальным полем разложения*, если $E = F(\alpha_1, \ldots, \alpha_n)$, где $\alpha_1, \ldots, \alpha_n$ — все корни многочлена $f(x)$.

Теорема 2.14. Пусть $\sigma : F \to E$ — изоморфизм полей, $f(x) \in F[x]$, $\overline{F}$ — некоторое минимальное поле разложения многочлена $f(x)$, $\overline{E}$ — некоторое минимальное поле разложения многочлена $\sigma(f(x)) \in E[x]$. Тогда существует изоморфизм $\tau : \overline{F} \to \overline{E}$, при котором $\tau|_F = \sigma$.

Доказательство. Проведем индукцию по $d_F(f)$, определенным в теореме 2.13. Пусть $d_F(f) = 0$. Тогда $\overline{F} = F$ и по предложению 2.45 $d_E(\sigma(f)) = 0$, поэтому $\overline{E} = E$.

Предположим, что теорема верна для любого многочлена $f(x) \in F[x]$ с условием $d_F(f) < k$. Покажем справедливость теоремы для случая $d_F(f) = k$. Предположим, что в разложении (2.6) $\deg\ p_1(x) > 1$. Понятно, что, учитывая предложение 2.45:

$$\sigma(f(x)) = \sigma(a_n)\sigma(p_1(x))^{k_1} \ldots \sigma(p_s(x))^{k_s}$$

— каноническое разложение многочлена $\sigma(f(x))$ над полем E. Пусть α — корень многочлена $p_1(x)$ в поле $\overline{F}$, β — корень многочлена $\sigma(p_1(x))$ в поле $\overline{E}$. Так как по предложению 2.46 поля $F(\alpha)$ и $E(\beta)$ изоморфны, причем $d_{F(\alpha)}(f) < k$, то доказательство следует из предположения индукции. □

Следствие 2.10. Пусть E_1 и E_2 — некоторые минимальные поля разложения многочлена $f(x) \in F[x]$. Тогда поля E_1 и E_2 изоморфны над F.

2.5.6. Конечные поля

Теорема 2.15. Если F — конечное поле, то $|F| = p^n$ для некоторого простого p и натурального n, причем:

(i) $char\ F = p$;

(ii) $n = [F : F_0]$, где F_0 — простое подполе поля F;

(iii) F является минимальным полем разложения многочлена $x^{p^n} - x \in F_0[x]$ и совпадает с множеством всех его корней.

Доказательство. В силу конечности поля F, конечно его простое подполе F_0 и конечна степень расширения $[F : F_0]$. По теореме 2.9 простое подполе F_0 изоморфно $\mathbb{Z}_p$ для некоторого простого p. Поэтому $char\ F = p$ и $|F| = p^n$ для некоторого $n = [F : F_0]$.

Покажем утверждение (iii). Так как порядок мультипликативной группы F^* равен числу $p^n - 1$, то по теореме Лагранжа $x^{p^n-1} = 1$ для любого $x \in F^*$. Поэтому $x^{p^n} = x$ для любого $x \in F$. □

Теорема 2.16. Для любого простого числа p и любого натурального n существует, и притом единственное с точностью до изоморфизма, поле, состоящее из p^n элементов.

Доказательство. Рассмотрим многочлен:

$$f(x) = x^{p^n} - x \in \mathbb{Z}_p[x].$$

По теореме 2.13 существуем минимальное поле разложения E многочлена $f(x)$. Поскольку:

$$f'(x) = p^n x^{p^n-1} - 1 = -1, \quad (f(x), f'(x)) = 1,$$

то многочлен $f(x)$ не имеет в поле E кратных корней. Пусть $K = \{\alpha_1, \ldots, \alpha_{p^n}\}$ — все корни многочлена $f(x)$ в поле E. Покажем замкнутость данного множество относительно операций сложения и умножения. Так как $char\ E = p$ и $\alpha_i^{p^n} = \alpha_i$, $i = 1, \ldots, p^n$, то:

$$(\alpha_i + \alpha_j)^{p^n} = \alpha_i^{p^n} + \alpha_j^{p^n} = \alpha_i + \alpha_j,$$

$$(\alpha_i \alpha_j)^{p^n} = \alpha_i^{p^n} \alpha_j^{p^n} = \alpha_i \alpha_j.$$

Поэтому по предложению 2.42 множество K является полем, а в силу минимальности поля E, $E = K$.

Далее, пусть F и E — некоторые поля с единичными элементами 1_F и 1_E, состоящие из p^n элементов. Пусть также F_0 и E_0 — соответствующие простые подполя данных полей, причем $F_0 \cong \mathbb{Z}_p \cong E_0$. По теореме 2.15 поле F является минимальным полем разложения многочлена $1_F x^{p^n} - 1_F x \in F_0[x]$. Аналогично,

поле E является минимальным полем разложения многочлена $1_E x^{p^n} - 1_E x \in E_0[x]$. По теореме 2.14 поля F и E изоморфны. $\square$

Следствие 2.11 (**теорема Вильсона**). Число p является простым тогда и только тогда, когда $(p-1)! \equiv -1 \pmod{p}$.

Доказательство. Пусть p — простое. Тогда все ненулевые элементы поля $\mathbb{Z}_p$ являются корнями многочлена $x^{p-1} - 1$. Поэтому:

$$x^{p-1} - 1 = (x-1)(x-2)\dots(x-(p-1)).$$

Приравнивая свободные члены данных многочленов, получаем $(-1)^{p-1}(p-1)! \equiv -1 \pmod{p}$.

Обратно, пусть $(p-1)! \equiv -1 \pmod{p}$. Предположим, что p — составное. Пусть $p = p_1^{\alpha_1} \dots p_n^{\alpha_n}$ — каноническое разложение числа p. Очевидно, что все множители данного разложения вида $p_i^{\alpha_i}$ попарно различны. Поэтому p является делителем числа $(p-1)!$. Противоречие. $\square$

Следствие 2.12 (**малая теорема Ферма**). Пусть p — простое и a — целое число, причем $(a,p) = 1$. Тогда $a^{p-1} \equiv 1 \pmod{p}$.

Доказательство. Так как $(a,p) = 1$, то a является корнем многочлена $x^{p-1} - 1 : a^{p-1} - 1 \equiv 0 \pmod{p}$. $\square$

Теорема 2.16 позволяет фиксировать произвольное поле из p^n элементов, которое обозначается через $GF(p^n)$.

Предложение 2.47. 1. Для любых $s, n \in \mathbb{N}$ и любого поля F многочлен $x^n - 1 \in F[x]$ делит многочлен $x^{sn} - 1 \in F[x]$.

2. Для любых натуральных d, n и любого поля F многочлен $x^n - x \in F[x]$ делит многочлен $x^{n^d} - x \in F[x]$.

Доказательство 1. Следует из равенства:

$$x^{sn} - 1 = (x^n - 1)(x^{(s-1)n} + x^{(s-2)n} + \dots + x^n + 1).$$

2. Следует из предыдущего пункта, так как:

$$x^n - x = x(x^{n-1} - 1),$$

$$x^{n^d} - x = x(x^{n^d-1} - 1) = x(x^{(n-1)s} - 1),$$

где $s = n^{d-1} + n^{d-2} + \ldots + n + 1$. □

Предложение 2.48. Для любых $s, n, a \in \mathbb{N}$ число $a^n - 1$ делит число $a^{sn} - 1$.

Доказательство следует из равенства:

$$a^{sn} - 1 = (a^n - 1)(a^{(s-1)n} + a^{(s-2)n} + \ldots + a^n + 1).$$

Предложение 2.49. Поле $GF(p^n)$ содержит подполе из r элементов тогда и только тогда, когда $r = p^s$ и s делит n. Это подполе единственно и изоморфно полю $GF(p^s)$.

Доказательство. Пусть $K \subset GF(p^n)$ подполе из r элементов. Рассмотрим $GF(p^n)$ как векторное пространство над K. Пусть $[GF(p^n) : K] = d$. Тогда $r^d = p^n$. Отсюда $r = p^s$, где $sd = n$.

Обратно, пусть $r = p^s$, где $sd = n$ для некоторого $d \in \mathbb{N}$. Пусть $K \subset GF(p^n)$ — подмножество, состоящее из всех корней многочлена $x^{p^s} - x$. По предложению 2.47 многочлен $x^{p^s} - x$ делит многочлен $x^{p^n} - x$. Поэтому все корни имеют кратность 1 и их ровно p^s. С другой стороны, как и в доказательстве теоремы 2.16 легко показать, что K — поле.

Предположим, что в $GF(p^n)$ содержится подполе E, содержащее p^s элементов. Если $K \neq E$, то многочлен $x^{p^s} - x$ имеет в поле $GF(p^n)$ более чем p^s корней, что невозможно. Значит, K — единственное подполе в $GF(p^n)$, содержащее p^s элементов. □

Следствие 2.13. В поле $GF(p^n)$ для любого $s \in \mathbb{N}$, $s|n$, существует единственное подполе из p^s элементов. Этими полями исчерпываются все подполя в $GF(p^n)$.

2.5.7. Образующие элементы конечного поля

Определение 2.27. Элемент g конечного поля $GF(q)$ порядка $q = p^n$ называется *образующим*, если порядок элемента g равен $q - 1$, т.е. $GF(q)^* = \{g, g^2, \ldots, g^{q-1} = 1\}$.

Теорема 2.17. Пусть $GF(q)$ — конечное поле порядка $q = p^n$, $q - 1 = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ — каноническое разложение числа $q - 1$. Элемент $g \in GF(q)$ является образующим тогда и только тогда, когда:

$$g^{\frac{q-1}{p_i}} \neq 1, \quad i = 1, \dots, k.$$

Доказательство аналогично доказательству теоремы 1.45.

Теорема 2.18. Каждое конечное поле $GF(q)$ порядка $q = p^n$ имеет образующий элемент. Более того, поле $GF(q)$ имеет ровно $\varphi(q-1)$ различных образующих элементов, где φ — функция Эйлера, причем если g — образующий элемент, то все данные $\varphi(q-1)$ образующих элемента составляют множество:

$$\{g^i \mid i = 1, \dots, q-2,\ (i, q-1) = 1\}.$$

Доказательство. Поскольку поле $GF(q)$ конечно, то по теореме Лагранжа каждый элемент в нем имеет конечный порядок, не превышающий числа $q-1$. Разобьем множество $GF(q)^*$ на непересекающиеся классы следующим образом. Скажем, что два элемента из $GF(q)^*$ принадлежат одному и тому же классу K_d, ели они имеют один и тот же порядок d. При этом из теоремы Лагранжа следует, что если элемент g имеет порядок d, то d делит $q-1$. Поэтому количество классов не превосходит количества делителей числа $q-1$ и верно равенство:

$$\sum_{d \mid (q-1)} |K_d| = q - 1. \tag{2.7}$$

Зафиксируем некоторое d, делящее $q-1$. Предположим, что множество K_d непусто (далее будет показано, что для любого делителя d числа $q-1$ множество K_d непусто) и $g \in K_d$. Следующая последовательность чисел состоит из попарно различных элементов:

$$g,\ g^2, \dots, g^d = 1,$$

которые образуют циклическую группу $\langle g \rangle$ порядка d. При этом $K_d \subseteq \langle g \rangle$, так как любой элемент из $GF(q)^*$ порядка d является

корнем уравнения $x^d = 1$, которое имеет не более d корней, причем все d элементов циклической группы $\langle g \rangle$ являются корнями данного уравнения.

Множество K_d совпадает со множеством всех тех элементов из $\langle g \rangle$, порядок которых равен d, а это все элементы, являющиеся образующими циклической группы $\langle g \rangle$. Из предложения 2.15 следует, что $|K_d| = \varphi(d)$.

Таким образом, для любого $d \mid (q-1)$ либо $|K_d| = 0$, либо $|K_d| = \varphi(d)$. Но из равенств (2.7) и:

$$\sum_{d \mid (q-1)} \varphi(d) = q - 1$$

(см. следствие 1.7) будет следовать, что $|K_d| = \varphi(d)$ для любого $d \mid (q-1)$. В частности, $|K_{q-1}| = \varphi(q-1)$, т.е. существует ровно $\varphi(q-1)$ элементов в $GF(q)^*$ порядка $q-1$. □

Следствие 2.14. Мультипликативная группа конечного поля является циклической.

Следствие 2.15. Для любого расширения конечных полей $F \subset E$ существует элемент $\alpha \in E$, который порождает E над F : $E = F(\alpha)$.

Доказательство. Пусть g — образующий элемент поля E. Тогда $E = F(g)$. □

Следствие 2.16. Для любого простого числа p найдется такое целое число g, что все его степени пробегают все ненулевые классы вычетов по модулю p.

Пример 2.26. Все ненулевые вычеты по модулю числа $p = 7$ можно получить из степеней числа 3: 3, 2, 6, 4, 5, 1.

2.5.8. Неприводимые многочлены над конечными полями

В следующей теореме показано, что над любым конечным полем существуют неприводимые многочлены любой степени $n \in \mathbb{N}$.

Теорема 2.19. Если $F = GF(p^d)$, то для любого $n \in \mathbb{N}$ существует многочлен из $F[x]$ степени n, неприводимый над F.

Доказательство. Пусть $E = GF(p^{nd})$, $F \subset E$ (теорема 2.16, следствие 2.13). По следствию 2.15 $E = F(g)$, где g — образующий элемент поля E.

Так как $|F| = p^d$, $|E| = p^{nd}$, то $[E : F] = [F(g) : F] = n$. Элемент g алгебраичен над полем F. Поэтому по теореме 2.11 $n = [E : F] = \deg\, Irr(g, F, x)$. □

Теорема 2.20. Пусть $f(x)$ — неприводимый многочлен степени n над полем $F = GF(q)$, $q = p^d$, и $E = F(\alpha)$ — расширение поля F, порожденное корнем α многочлена $f(x)$. Тогда выполнены следующие утверждения:

1) E — минимальное поле разложения многочлена $f(x)$ над полем F, причем элементы:

$$\alpha,\ \alpha^q,\ \alpha^{q^2}, \ldots, \alpha^{q^{n-1}} \tag{2.8}$$

являются попарно различными n корнями многочлена $f(x)$;

2) $f(x) \mid x^{q^n} - x$.

Доказательство. 1. Так как $f(x)$ неприводим над F и $\alpha \in E$ — корень многочлена $f(x)$, то α — алгебраический элемент поля E над F, причем из предложения 2.43 и теоремы 2.11 следует, что:

$$E = \langle 1,\ \alpha,\ \alpha^2, \ldots,\ \alpha^{n-1} \rangle_F, \quad n = \deg\ f(x) = [E : F], \quad |E| = q^n$$

и любой элемент $b \in E$ имеет вид $b = \sum_{i=0}^{n-1} b_i \alpha^i$, $b_i \in F$, $i = 0, \ldots, n-1$.

Из теоремы 2.15 следует, что $a^{q^s} = a$ для любых $a \in F$, $s \in \mathbb{N}$. Действительно, из теоремы 2.15 следует, что $a^q = a$. Равенство $a^{q^s} = (a^{q^{s-1}})^q = a$ следует из предположения и базы индукции по s.

Пусть $f(x) = \sum_{i=0}^{n-1} f_i x^i$. В связи с тем, что $char\ F = p$, выполнено равенство:

$$f(\alpha^{q^s}) = \sum_{i=0}^{n-1} f_i (\alpha^i)^{q^s} = \sum_{i=0}^{n-1} (f_i \alpha^i)^{q^s} = \left(\sum_{i=0}^{n-1} f_i \alpha^i \right)^{q^s} = f(\alpha)^{q^s} = 0.$$

Поэтому все элементы вида (2.8) являются корнями $f(x)$. Покажем, что данные элементы попарно различны. Допустим, что $\alpha^{q^s} = \alpha^{q^t}$, $0 \leqslant s < t \leqslant n-1$. Тогда при $r = t - s$:

$$0 = \alpha^{q^t} - \alpha^{q^s} = \alpha^{q^{r+s}} - \alpha^{q^s} = (\alpha^{q^r})^{q^s} - \alpha^{q^s} = (\alpha^{q^r} - \alpha)^{q^s}.$$

Из этого следует, что $\alpha^{q^r} = \alpha$, $0 < r < n$.

Пусть $E \ni b = \sum_{i=0}^{n-1} b_i \alpha^i$. Так как $b_i^{q^r} = b_i$ для любого $i = 0, \ldots, n-1$, то:

$$b^{q^r} = \left(\sum_{i=0}^{n-1} b_i \alpha^i\right)^{q^r} = \sum_{i=0}^{n-1} (b_i \alpha^i)^{q^r} = \sum_{i=0}^{n-1} b_i^{q^r} (\alpha^{q^r})^i = \sum_{i=0}^{n-1} b_i \alpha^i = b.$$

Следовательно, все q^n элементов поля E являются корнями многочлена $x^{q^r} - x$, что невозможно в силу условия $r < n$. Поэтому все элементы вида (2.8) попарно различны.

2. По теореме 2.15 все элементы поля E — корни многочлена $x^{q^n} - x$. Осталось применить предложение 2.43. □

Пример 2.27. Построим поле $GF(4)$ из 4 элементов. Согласно теореме 2.16 это можно сделать. Согласно теореме 2.11 и следствию 2.15 $GF(4) = \langle 1, \alpha \rangle_{\mathbb{Z}_2}$ для некоторого элемента $\alpha \in GF(4)$. Для составления таблицы умножения в данном поле необходимо найти минимальный многочлен $Irr(\alpha, \mathbb{Z}_2, x)$ степени 2. Такой многочлен существует (теорема 2.19). Из теоремы 2.20 следует, что такой многочлен следует искать среди делителей многочлена $x^4 - x$:

$$x^4 - x = x(x^3 - 1) = x(x-1)(x^2 + x + 1).$$

Многочлен $x^2 + x + 1$ неприводим над полем $\mathbb{Z}_2$. В этом случае $\alpha^2 = -\alpha - 1 = \alpha + 1$. Таблицы сложения и умножения будут иметь такой вид:

$+$	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	1	0	$\alpha+1$	α
α	α	$\alpha+1$	0	1
$\alpha+1$	$\alpha+1$	α	1	0

$\cdot$	0	1	α	$\alpha+1$
0	0	0	0	0
1	0	1	α	$\alpha+1$
α	0	α	$\alpha+1$	1
$\alpha+1$	0	$\alpha+1$	1	α

Пример 2.28. Построим $GF^*(2^4)$ по модулю многочлена $p(x) = x^4 + x + 1$, при условии $p(\alpha) = 0 : \alpha^4 = \alpha + 1$. Имеем:

$$
\begin{array}{rrrrrcl}
\alpha^0 = & & & & 1 & = & 0001, \\
\alpha^1 = & & & \alpha & & = & 0010, \\
\alpha^2 = & & \alpha^2 & & & = & 0100, \\
\alpha^3 = & \alpha^3 & & & & = & 1000, \\
\alpha^4 = & & & \alpha & +1 & = & 0011, \\
\alpha^5 = & & \alpha^2 & +\alpha & & = & 0110, \\
\alpha^6 = & \alpha^3 & +\alpha^2 & & & = & 1100, \\
\alpha^7 = & \alpha^3 & & +\alpha & +1 & = & 1011, \\
\alpha^8 = & & \alpha^2 & & +1 & = & 0101, \\
\alpha^9 = & \alpha^3 & & +\alpha & & = & 1010, \\
\alpha^{10} = & & \alpha^2 & +\alpha & +1 & = & 0111, \\
\alpha^{11} = & \alpha^3 & +\alpha^2 & +\alpha & & = & 1110, \\
\alpha^{12} = & \alpha^3 & +\alpha^2 & +\alpha & +1 & = & 1111, \\
\alpha^{13} = & \alpha^3 & +\alpha^2 & & +1 & = & 1101, \\
\alpha^{14} = & \alpha^3 & & & +1 & = & 1001, \\
\alpha^{15} = & & & & 1 & = & 0001.
\end{array}
\tag{2.9}
$$

Поэтому элемент α является образующим элементом поля $GF(2^4)$ по модулю многочлена $p(x)$. Из теоремы 2.18 следует, что элементы:

$$\alpha,\ \alpha^2,\ \alpha^4,\ \alpha^7,\ \alpha^8,\ \alpha^{11},\ \alpha^{13},\ \alpha^{14}$$

— все образующие поля $GF(2^4)$.

Соотношения (2.9) демонстрируют связь между аддитивной и мультипликативной группами поля. Например, нужно перемножить элементы поля $\alpha^2+\alpha$ и $\alpha^3+\alpha+1$, которые представимы соответственно векторами 0110 и 1011 соответственно. Учитывая (2.9), получаем:

$$(\alpha^2 + \alpha) \cdot (\alpha^3 + \alpha + 1) = \alpha^5 \cdot \alpha^7 = \alpha^{12} = \alpha^3 + \alpha^2 + \alpha + 1.$$

Обратно, пусть требуется сложить два элемента поля α^8 и α^{10}. Тогда:

$$\alpha^8 + \alpha^{10} = (\alpha^2 + 1) + (\alpha^2 + \alpha + 1).$$

Последняя сумма соответствует сумме $0101+0111 = 0010$, что соответствует элементу α.

2.5.9. Автоморфизм Фробениуса. Совершенные поля

Лемма 2.1. Пусть F — поле характеристики $char\ F = p > 0$. Тогда отображение $f : F \to F$, определенное правилом $f(x) = x^p$, $x \in F$, является изоморфизмом на некоторое подполе.

Доказательство. Гомоморфизм отображения f следует из соотношений:

$$(x + y)^p = x^p + y^p, \quad (xy)^p = x^p y^p.$$

Так как F — поле, то из равенства $x^p = 0$ следует, что $x = 0$, поэтому $Ker\ f = \{0\}$, что означает инъективность отображения f. □

Определенное выше отображение f называется *отображением Фробениуса*. Неподвижными элементами f являются корни уравнения $x^p = x$, а это — в точности элементы простого поля $\mathbb{Z}_p \subset F$.

Лемма 2.2. Пусть F – поле характеристики $char\ F = p > 0$ и $n \in \mathbb{N}$. Тогда множество $F^{p^n} = \{x^{p^n} \mid x \in F\}$ является подполем в F изоморфным полю F.

Доказательство. Из леммы 2.1 следует, что поля F и $f(F)$ изоморфны. Нетрудно видеть, что композиция $f^n = \underbrace{f \circ \ldots \circ f}_{n}$ является изоморфизмом полей F и $f^n(F)$, причем $f^n(F) = F^{p^n}$. □

Определение 2.28. Поле F называется *совершенным*, если $char\ F = 0$ или $char\ F = p > 0$ и $F^p = F$, т.е. отображение Фробениуса сюрьективно.

В совершенном поле положительной характеристики отображение Фробениуса является автоморфизмом.

Предложение 2.50. Любое конечное поле совершенно.

Доказательство следует из того факта, что отображение $\varphi : X \to Y$ конечных равномощных множеств X и Y инъективно тогда и только тогда, когда φ сюрьективно. $\square$

Пример 2.29. Пусть F — поле характеристики $char\ F = p > 0$ и $F(t)$ — поле рациональных дробей над F. Так как уравнение $x^p = t$ не имеет корней в $F(t)$, то поле $F(t)$ не является совершенным.

Предложение 2.51. Пусть K — совершенное поле и пусть F — его алгебраическое расширение. Тогда поле F также является совершенным.

Доказательство. Так как $K = K^p \subset F^p \subset F$, то поле F^p можно рассматривать как векторное пространство над полем K. Сначала рассмотрим случай, когда степень расширения $[F : K] = n$ конечна. В этом случае $F = \langle e_1, \ldots, e_n \rangle_K$, где $e_1, \ldots, e_n$ — базис векторного пространства F над K. Тогда элементы $e_1^p, \ldots, e_n^p$ поля F^p линейно независимы над K. Действительно, рассмотрим линейное соотношение $\sum_{i=1}^n \alpha_i e_i^p = 0$, $\alpha_i \in K$, $i = 1, \ldots, n$. Так как поле K совершенно, то для некоторых $\beta_i \in K$ выполнено равенство $\beta_i^p = \alpha_i$, $i = 1, \ldots, n$. Поэтому:

$$0 = \sum_{i=1}^n \alpha_i e_i^p = \sum_{i=1}^n \beta_i^p e_i^p = \left(\sum_{i=1}^n \beta_i e_i \right)^p, \quad \sum_{i=1}^n \beta_i e_i = 0.$$

Так как $\beta_i = 0$, то $\alpha_i = 0$, $i = 1, \ldots, n$. Таким образом, $\dim\ F^p \geqslant \dim\ F = n$. С учетом того, что $F^p \subset F$ получаем $F^p = F$.

Пусть теперь степень расширения $[F : K]$ не является конечной. Пусть $\alpha \in F$. В силу конечности степени расширения $[K(\alpha) : K]$ (теорема 2.11) из предыдущего случая следует, что $K(\alpha)$ — совершенное поле, поэтому уравнение $x^p - \alpha = 0$ имеет решение в поле F. $\square$

2.5.10. Трансцендентные расширения полей

Определение 2.29. Пусть F — расширение поля K. Подмножество $M \subset F$ называется *алгебраически независимым* над

K, если для любого многочлена $f(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n]$ из соотношения:

$$f(a_1, \ldots, a_n) = 0, \quad a_i \in M, \quad a_i \neq a_j,$$

следует, что $f(x_1, \ldots, x_n) = 0$. Подмножество $M \subset F$ называется *базисом трансцендентности*, если выполнены следующие условия:

1) M алгебраически независимо;

2) любое алгебраически независимое подмножество $\widetilde{M} \subset F$, содержащее M, совпадает с M.

Предложение 2.52. Алгебраически независимое подмножество M является базисом трансцендентности F над K тогда и только тогда, когда поле F алгебраично над $K(M)$.

Доказательство. Пусть M — базис трансцендентности F над K и $a \in F \setminus M$. Тогда найдутся такие $a_1, \ldots, a_n \in M$, $a_j \neq a_j, i \neq j$, что $f(a_1, \ldots, a_n, a) = 0$ для некоторого ненулевого многочлена $f(x_1, \ldots, x_{n+1}) \in K[x_1, \ldots, x_{n+1}]$. Так как:

$$0 = f(a_1, \ldots, a_n, a) = g_0 + g_1 a + \ldots + g_m a^m, \quad g_i = g_i(a_1, \ldots, a_n),$$

$$g_i(x_1, \ldots, x_n) \in K[x_1, \ldots, x_n], \quad i = 0, \ldots, m,$$

то a алгебраичен над $K(M)$.

Обратно, пусть поле F алгебраично над $K(M)$ и $a \in F \setminus M$. Найдется такой ненулевой многочлен $\sum_{i=0}^{m} g_i x^i \in K(M)[x]$. корнем которого является a. Домножим данный многочлен на наименьшее общее кратное знаменателей коэффициентов g_0,. . . ,g_m. Тогда полученный многочлен принадлежит кольцу$K[M][x]$, корнем которого является a. □

Пример 2.30. 1. Множество из одного элемента $\alpha \in F$ алгебраически независимо над K тогда и только тогда, когда α трансцендентен над K.

2. Пусть $F = K(x_1, \ldots, x_n)$ — поле рациональных дробей от n переменных. Тогда переменные x_1,. . . , x_n алгебраически независимы и образуют базис трансцендентности. Верно и обратное: если $F = K(a_1, \ldots, a_n)$, где элементы a_1,. . . , a_n образуют базис

трансцендентности, то F изоморфно полю рациональных дробей от n переменных.

Определение 2.30. Пусть расширение $K \subset F$ конечно порождено, т.е. $F = K(\alpha_1, \ldots, \alpha_n)$ для некоторых $\alpha_1, \ldots, \alpha_n \in F$. Если M — алгебраически независимое подмножество в F и если мощность множества M является наибольшей среди мощностей всех таких подмножеств, то будем называть эту мощность *степенью трансцендентности* расширения F над K и обозначать $degtr_K\ F$ или просто $degtr\ F$.

Лемма 2.3. Пусть F — расширения поля K. Если множество $A = \{\alpha_1, \ldots, \alpha_n\}$ порождает F над K (т.е. $F = K(A)$) и M — подмножество в A, алгебраически независимое над K, то существует базис трансцендентности S поля F над K такой, что $M \subseteq S \subseteq A$.

Доказательство. Можно считать, что $M = \{\alpha_1, \ldots, \alpha_m\}$, $m \leqslant n$, и каждая система $\alpha_1, \ldots, \alpha_m, \alpha_j$, $j = m+1, \ldots, n$, алгебраически зависима. Тогда элементы $\alpha_{m+1}, \ldots, \alpha_n$ алгебраичны над $K(\alpha_1, \ldots, \alpha_m)$. В этом случае $S = M$, так как каждый этаж башни расширений:

$$K(S) \subset K(S, \alpha_{m+1}) \subset \ldots \subset K(S, \alpha_{m+1}, \ldots, \alpha_n) = F$$

конечен (теорема 2.11), то конечно и расширение $K(S) \subset F$ (теорема 2.10), а, следовательно, алгебраично (предложение 2.44). Осталось применить предложение 2.52. □

Лемма 2.4 (об алгебраической зависимости)**.** Пусть F — расширение поля K и пусть $\alpha_1, \ldots, \alpha_n$ — базис трансцендентности поля F над K. Если $\beta_1, \ldots, \beta_m$ — элементы поля F, алгебраически независимые над K, то $m \leqslant n$.

Доказательство. Можно считать, что $\alpha_1, \ldots, \alpha_n$ — базис трансцендентности из минимального числа элементов.

Индукцией по i покажем, что для любого $i = 1, \ldots, m$ элементы $\beta_1, \ldots, \beta_i, \alpha_{i+1}, \ldots, \alpha_n$ образуют базис трансцендентности.

Элементы β_1, $\alpha_1,\ldots$, α_n алгебраически зависимы, поэтому найдется такой ненулевой многочлен:

$$f(x_1,\ldots,x_{n+1}) \in K[x_1,\ldots,x_{n+1}],$$

что $f(\beta_1,\alpha_1,\ldots,\alpha_n) = 0$, причем в многочлене f встречается переменная x_1 и некоторая переменная x_i, $2 \leqslant i \leqslant n+1$, скажем x_2, также встречается в f. Тогда элемент α_1 алгебраичен над $K(\beta_1,\alpha_2,\ldots,\alpha_n)$. Так как поле F алгебраично над $K(\alpha_1,\ldots,\alpha_n)$, то F также алгебраично над $K(\beta_1,\alpha_1,\ldots,\alpha_n)$. Поэтому из алгебраичности каждого этажа башни расширений:

$$K(\beta_1,\alpha_2,\ldots,\alpha_n) \subset K(\beta_1,\alpha_1,\ldots,\alpha_n) \subset F$$

следует, что поле F алгебраично над $K(\beta_1,\alpha_2,\ldots,\alpha_n)$ (следствие 2.9). Заметим, что элементы β_1, $\alpha_2,\ldots,\alpha_n$ алгебраически независимы, так как $\alpha_1,\ldots$, α_n — базис трансцендентности из минимального числа элементов. Из предложения 2.52 следует, что β_1, $\alpha_2,\ldots,\alpha_n$ — базис трансцендентности F над K.

Предположим, что $\beta_1,\ldots,\beta_{i-1}$, $\alpha_i,\ldots$, α_n — базис трансцендентности, $i \geqslant 2$. Как и выше, элементы $\beta_1,\ldots$, β_i, $\alpha_i,\ldots$, α_n алгебраически зависимы, поэтому найдется такой ненулевой многочлен $g(x_1,\ldots,x_{n+1}) \in K[x_1,\ldots,x_{n+1}]$, что:

$$g(\beta_1,\ldots,\beta_i,\alpha_i,\ldots,\alpha_n) = 0,$$

причем в многочлене g встречается переменная x_i и некоторая переменная x_j, $i+1 \leqslant j \leqslant n+1$, скажем x_{i+1}, также встречается в g. Тогда элемент α_i алгебраичен над $K(\beta_1,\ldots,\beta_i,\alpha_{i+1},\ldots,\alpha_n)$. По предположению индукции каждый этаж башни расширений:

$$K(\beta_1,\ldots,\beta_i,\alpha_{i+1},\ldots,\alpha_n) \subset K(\beta_1,\ldots,\beta_i,\alpha_i,\ldots,\alpha_n) \subset F$$

алгебраичен, поэтому поле F алгебраично над:

$$K(\beta_1,\ldots,\beta_i,\alpha_{i+1},\ldots,\alpha_n).$$

С учетом минимальности n $\beta_1,\ldots,\beta_i$, $\alpha_{i+1},\ldots$, α_n — базис трансцендентности.

Исходя из доказанного, если $m > n$, то элемент β_{n+1} должен быть алгебраичным над $K(\beta_1, \ldots, \beta_n)$, что привело бы к противоречию с тем, что $\beta_1, \ldots, \beta_m$ — алгебраически независимы над K. □

Теорема 2.21. Пусть расширение $K \subset F$ конечно порождено. Тогда любые два базиса трансцендентности поля F над K имеют одинаковую мощность.

Доказательство следует из леммы 2.4. □

Глава 3. Элементы алгебраической геометрии

3.1. Аффинные алгебраические многообразия

Пусть F — некоторое алгебраически замкнутое поле (поле, в котором всякий многочлен ненулевой степени над F имеет хотя бы один корень). *Аффинным n-мерным пространством* $\mathbf{A}^n = \mathbf{A}^n(F)$ над полем F называется следующее множество наборов из всех n элементов поля F :

$$\mathbf{A}^n(F) = \{(a_1, \ldots, a_n) \in F^n\}.$$

Упорядоченный набор $P = (a_1, \ldots, a_n) \in \mathbf{A}^n$ называется точкой пространства $\mathbf{A}^n$, а его компоненты $a_1, \ldots, a_n$ — *аффинными координатами* точки P. При $n = 1$ аффинное пространство называется *аффинной прямой*, при $n = 2$ — *аффинной плоскостью*.

Элементы кольца многочленов $F[x_1, \ldots, x_n]$ будем интерпретировать как функции на n-мерном аффинном пространстве со значениями в F, полагая $f(P) = f(a_1, \ldots, a_n)$, где $f \in F[x_1, \ldots, x_n]$, $P \in \mathbf{A}^n$. Рассмотрим *множество нулей* произвольного многочлена $f \in F[x_1, \ldots, x_n]$:

$$Z(f) = \{P \in \mathbf{A}^n \mid f(P) = 0\}.$$

Пусть T — некоторое подмножество в $F[x_1, \ldots, x_n]$. Аналогично, можно определить множество нулей подмножества T :

$$Z(T) = \{P \in \mathbf{A}^n \mid f(P) = 0 \quad \forall f \in T\}.$$

Заметим, что если I — идеал в $F[x_1, \ldots, x_n]$, порожденный подмножеством T, то $Z(I) = Z(T)$. Действительно, идеал I состоит из всех конечных сумм вида $\sum_{i=1}^{s} f_i g_i$, $s \in \mathbb{N}$, $f_i \in T$,

$g_i \in F[x_1, \ldots, x_n]$. Осталось заметить, что для некоторой точки $P \in \mathbf{A}^n$ выполнены равенства $f_i(P) = 0$, $i = 1, \ldots, s$, тогда и только тогда, когда $h(P) = 0$ для любого $h \in I$.

Определение 3.1. Подмножество X в $\mathbf{A}^n$ называется (*аффинным*) *алгебраическим множеством*, если существует такое подмножество $T \subseteq F[x_1, \ldots, x_n]$, что $X = Z(T)$. Непустое алгебраическое множество называется также *аффинным алгебраическим многообразием* или просто *аффинным многообразием*.

Замечание 3.1. Отметим, что каждое подмножество $T \subseteq F[x_1, \ldots, x_n]$ однозначно определяет подмножество $Z(T)$. В то же время, T не может быть однозначно определено по $Z(T)$. Например, многочлены f и f^k (k — любое натуральное число) определяют одно и то же подмножество: $Z(f) = Z(f^k)$.

При этом из сказанного выше следует, что алгебраическое множество X однозначно определяется идеалом:

$$I(X) = \{f \in F[x_1, \ldots, x_n] \mid f(P) = 0 \quad \forall P \in X\}.$$

Определение 3.2. Идеал $I(X)$ называется *идеалом алгебраического множества* X.

Будем говорить, что алгебраическое многообразие $X \subseteq \mathbf{A}^n$ является *гиперповерхностью*, заданной многочленом f, если f порождает идеал $I(X) : I(X) = (f)$.

Идеал $I(X)$ всегда конечно порожден, что следует из теоремы 3.1.

Теорема 3.1 (Гильберт). Каждый идеал кольца многочленов $F[x_1, \ldots, x_n]$ конечно порожден.

Это означает, что всегда в качестве множества T можно рассматривать только конечные множества.

Пример 3.1 (алгебраические многообразия и их идеалы). 1. $Z(0) = \mathbf{A}^n$, т.е. многообразие, порожденное нулевым полиномом, совпадает с аффинным пространством $\mathbf{A}^n$.

2. Пусть $a, b, c \in F$, $a^2 + b^2 > 0$ и $n = 2$. Тогда многообразие, порожденное полиномом $ax + by + c$, будет являться прямой на аффинной плоскости.

3. Пусть полиномы $f_1, \ldots, f_n$ линейны по каждой из переменных $x_1, \ldots, x_n$ и являются линейно независимыми. Тогда многообразие, порожденное данными полиномами, состоит из одной точки $(0, \ldots, 0) \in F^n$.

4. Любой полином нулевой степени порождает алгебраическое множество, равное пустому множеству.

5. Выясним как устроены алгебраические множества на аффинной прямой $\mathbf{A}^1$. Каждый идеал в кольце $F[x]$ является главным, поэтому каждое алгебраическое множество — это множество нулей одного многочлена. Так как поле F алгебраически замкнуто, то всякий ненулевой многочлен может быть записан в виде $f(x) = c(x - a_1)^{k_1} \ldots (x - a_s)^{k_s}$, где $a_1, \ldots, a_s \in F$. В этом случае $Z(f) = \{a_1, \ldots, a_s\}$. Таким образом, алгебраические множества в $\mathbf{A}^1$ — это пустое множество, всевозможные конечные подмножества и вся аффинная прямая.

Предложение 3.1. Пусть X_1, X_2 — алгебраические множества.

1. $X_1 \subseteq X_2$ тогда и только тогда, когда $I(X_2) \subseteq I(X_1)$.
2. $X_1 = X_2$ тогда и только тогда, когда $I(X_1) = I(X_2)$.
3. $X_1 \subset X_2$ тогда и только тогда, когда $I(X_2) \subset I(X_1)$.

Доказательство. 1. Пусть $X_1 \subseteq X_2$, $f \in I(X_2)$. Исходя из определения $I(X_2)$, $f(P) = 0$ для любого $P \in X_2$. Так как $X_1 \subseteq X_2$, то $f(P) = 0$ для любого $P \in X_1$. Поэтому $f \in I(X_1)$.

Обратно, пусть $I(X_2) \subseteq I(X_1)$. Так как для любого $f \in I(X_1)$ выполнено $f(X_1) = 0$, то $f(X_2) = 0$ для любого $f \in I(X_2)$. Поэтому $X_1 \subseteq X_2$.

2. Следует из замечания 3.1.

3. Следует из пунктов 1 и 2. □

Предложение 3.2. 1. Пересечение любого семейства алгебраических множеств также будет алгебраическим множеством.

2. Объединение любого конечного числа алгебраических множеств является алгебраическим множеством.

3. Пустое множество и все аффинное пространство являются алгебраическими множествами.

Доказательство. 1. Пусть $X_\alpha = Z(T_\alpha)$ — произвольное семейство алгебраических множеств. Тогда $\cap X_\alpha = Z(\cup T_\alpha)$.

2. Если $X_i = Z(T_i)$, $i = 1, \ldots, s$, то $X_1 \cup \ldots \cup X_s = Z(T_1 \ldots T_s)$, где $T_1 \ldots T_s$ — множество всех конечных сумм вида:

$$\sum f_1 \cdot \ldots \cdot f_s, \quad f_i \in T_i, \ i = 1, \ldots, \ s.$$

3. Пустое множество представляется в виде $\varnothing = Z(1)$, а все аффинное пространство — в виде $\mathbf{A}^n = Z(0)$. $\square$

Неприводимые аффинные многообразия.

Определение 3.3. Аффинное алгебраическое многообразие $X \subseteq \mathbf{A}^n$ называется *неприводимым*, если его нельзя представить в виде объединения $X = X_1 \cup X_2$ двух собственных алгебраических многообразий.

Пример 3.2. 1. Аффинная прямая $\mathbf{A}^1$ неприводима, так как ее собственные алгебраические подмножества конечны, а $\mathbf{A}^1$ — бесконечное множество (поле F алгебраически замкнуто, поэтому бесконечно).

2. Любое аффинное многообразие $X \subseteq \mathbf{A}^n$ содержит неприводимое подмногообразие — точку. При этом заметим, что любая аффинная точка $P = (a_1, \ldots, a_n) \in \mathbf{A}^n$ — это аффинное многообразие, так как $P = Z(x_1 - a_1, \ldots, \ x_n - a_n)$.

Предложение 3.3. Аффинное многообразие $X \subseteq \mathbf{A}^n$ неприводимо тогда и только тогда, когда его идеал $I(X)$ прост.

Доказательство. Пусть X неприводимо. Учитывая предложение 2.39, предположим, что $fg \in I(X)$, но $f, g \notin I(X)$. Обозначим:

$$I = I(X), \ I_1 = (I, f) = I + fF[x_1, \ldots, x_n],$$

$$I_2 = (I, g) = I + gF[x_1, \ldots, x_n].$$

Понятно, что:

$$I \subset I_1, \ I \subset I_2, \ Z(I_1) \subset X, \ Z(I_2) \subset X, \ I_1 I_2 \subseteq I.$$

Из предложения 3.1 следует, что:

$$X = Z(I) \subseteq Z(I_1 I_2) = Z(I_1) \cup Z(I_2).$$

С другой стороны, понятно, что $Z(I_1) \cup Z(I_2) \subseteq X$. Поэтому $X = Z(I_1) \cup Z(I_2)$. Противоречие с неприводимостью X.

Обратно, пусть идеал $I(X)$ прост. Предположим, что $X = X_1 \cup X_2$, где X_1, X_2 — алгебраические многообразия, причем $X_1 \subset X$, $X_2 \subset X$. Тогда из предложения 3.1 следует, что $I(X) \subset I(X_1)$, $I(X) \subset I(X_2)$. Пусть $f \in I(X_1) \setminus I(X)$, $g \in I(X_2) \setminus I(X)$. Так как $I(X) = I(X_1)I(X_2)$ (предложение 3.2), то $fg \in I(X)$. Противоречие. □

Следствие 3.1. Гиперповерхность $X \subseteq \mathbf{A}^n$, заданная многочленом f, неприводимо тогда и только тогда, когда многочлен f неприводим.

Следствие 3.2. Аффинное многообразие $\mathbf{A}^n$ неприводимо, так как $I(\mathbf{A}^n) = (0)$.

Теорема 3.2. Рассмотрим два многочлена $f, g \in F[x, y]$, где поле F произвольно. Если f неприводим, а система уравнений

$$\begin{cases} f(x, y) = 0, \\ g(x, y) = 0 \end{cases} \tag{3.1}$$

имеет бесконечное множество решений, то f делит g.

Доказательство. Можно считать, что многочлен f зависит от y. Рассмотрим f и g как многочлены над полем рациональных функций $F(x)$. Покажем, что f неприводим как элемент кольца $F(x)[y]$. Предположим, что это не так. Тогда существует нетривиальное разложение $f = f_1 f_2$, где $f_1, f_2 \in F(x)[y]$, причем f_1 и f_2 зависят от y. Пусть $\alpha(x)$ — общее кратное знаменателей коэффициентов f_1 и f_2. Тогда $\alpha(x)^2 f(x, y) = \widetilde{f_1}(x, y)\widetilde{f_2}(x, y)$, где $\widetilde{f_i} = \alpha(x) f_i(x, y) \in F[x, y]$. Так как $F[x, y]$ — кольцо с однозначным разложением на множители, то один из многочленов $\widetilde{f_1}$ или $\widetilde{f_2}$ делится на f. Предположим, что $\widetilde{f_1} = fh$, $h \in F[x, y]$. Тогда $\alpha(x)^2 f = fh\widetilde{f_2}$, из чего следует, что $\alpha(x)^2 = h\widetilde{f_2}$. Из этого равенства следует, что многочлен $\widetilde{f_2}$ не зависит от переменной y. Противоречие.

Итак, f неприводим как элемент кольца $F(x)[y]$. Предположим, что f не делит g. Нетрудно видеть, что f также не делит g в кольце $F(x)[y]$. Поэтому f и g взаимно просты в кольце $F(x)[y]$. Поэтому существуют $u, v \in F(x)[y]$ такие, что $uf + vg = 1$. Пусть $\beta(x)$ — общее кратное знаменателей коэффициентов u и v. Домножив обе части равенства $uf + vg = 1$ на $\beta(x)$, получим равенство:

$$\widetilde{u}(x,y)f(x,y) + \widetilde{v}(x,y)g(x,y) = \beta(x), \quad \widetilde{u} = \beta u,\ \widetilde{v} = \beta v.$$

Пусть $\{(a_n, b_n)\}_{n\geqslant 1}$ — бесконечная последовательность решений системы (3.1). Тогда $\{a_n\}_{n\geqslant 1}$ — бесконечная последовательность решений уравнения $\beta(x) = 0$, которое имеет лишь конечное множество решений. Поэтому последовательность $\{a_n\}_{n\geqslant 1}$ стационарная, начиная с некоторого номера $N : a_n = a$ для любого $n \geqslant N$. Из этого следует, что $\{b_n\}_{n\geqslant N}$ — последовательность решений уравнения $f(a, y) = 0$. Так как $f(x, y)$ не делится на $x - a$ (f неприводим), то многочлен $f(a, y)$ отличен от нуля и уравнение $f(a, y) = 0$ имеет конечное число решений. Противоречие с тем, что множество решений уравнения $f(x, y) = 0$ бесконечно. □

Следствие 3.3. Пусть F — произвольное поле, $f \in F[x, y]$ — неприводимый многочлен, $g \in F[x, y]$ — произвольный многочлен. Если g не делится на f, то система уравнений (3.1) имеет лишь конечное число решений.

Теорема 3.3. Пусть I — простой идеал кольца многочленов $F[x, y]$. Тогда будет верным одно из следующих равенств:

1) $I = (0)$;

2) $I = (f)$, где $f = f(x, y)$ — неприводимый многочлен;

3) $I = (x - a,\ y - b) = (x - a)F[x, y] + (y - b)F[x, y]$ для некоторых $a, b \in F$.

Доказательство. Предположим, что I не является главным идеалом. В силу простоты идеала I найдутся непропорциональные неприводимые многочлены $f, g \in I$. Для данных многочленов возможны следующие случаи:

1. f и g зависят от переменных x и y. Как и в доказательстве предыдущей теоремы, рассмотрим их как многочлены из $F(x)[y]$. В этом кольце многочленов f и g снова неприводимы и взаимно просты. Поэтому найдутся такие $u, v \in F(x)[y]$, что $uf + vg = 1$. Пусть $\alpha(x)$ — общее кратное знаменателей коэффициентов u и v. Домножая обе части равенства на $\alpha(x)$, получим $\widetilde{u}f + \widetilde{v}g = \alpha(x)$, где $\widetilde{u}, \widetilde{v} \in F[x, y]$. Так как $f, g \in I$, то $\alpha(x) \in I$. Поскольку поле F алгебраически замкнуто, то $\alpha(x) = c(x-a_1)\dots(x-a_s)$. В силу простоты идеала I получаем, что $x - a \in I$ для некоторого $a = a_i$. Аналогично, рассматривая f и g как элементы кольца $F(y)[x]$, получаем, что $y - b \in I$ для некоторого $b \in F$.

2. f зависит от x, g зависит от x и y. Так как поле F алгебраически замкнуто, то $f(x) = c(x - a_1)\dots(x - a_s)$ и $x - a \in I$ для некоторого $a = a_i$. Рассматривая f и g над $F(y)[x]$, получим, что $y - b \in I$ для некоторого $b \in F$.

3. f зависит от y, g зависит от x и y. Доказательство аналогично предыдущему случаю.

4. f зависит от x, g зависит от y. Тогда:

$$f(x) = c(x - a_1)\dots(x - a_s), \quad g(y) = d(y - b_1)\dots(y - b_t)$$

и $x - a \in I$ для некоторого $a = a_i$, $y - b \in I$ для некоторого $b = b_j$.

5. f и g зависят от переменной x либо f и g зависят от y. Тогда из алгоритма Евклида следует, что $(f, g) = 1$ и $I = F[x, y] = (1)$. Поэтому этот случай невозможен в силу того, что по предположению I не является главным идеалом.

Итак, для некоторых $a, b \in F$ $x - a \in I$, $y - b \in I$, поэтому $(x - a)F[x, y] + (y - b)F[x, y] \subseteq I$. При этом заметим, что для любого $\widetilde{a} \in F$, $\widetilde{a} \neq a$, следует, что $x - \widetilde{a} \notin I$, так как в противном случае $\widetilde{a} - a = (x - a) - (x - \widetilde{a}) \in I$, поэтому $1 \in I$ и $I = F[x, y]$. Аналогично, для любого $\widetilde{b} \in F$, $\widetilde{b} \neq b$, следует, что $y - \widetilde{b} \notin I$.

Пусть $f(x, y) = \sum \alpha_{ij} x^i y^j \in I$, $\alpha_{ij} \in F$. Многочлен f можно представить в виде:

$$f(x, y) = \sum \alpha_{ij}(x - a + a)^i (y - b + b)^j =$$

$$\sum_{i+j\geqslant 1} \beta_{ij}(x-a)^i(y-b)^j + c, \quad \beta_{ij}, c \in F.$$

Так как $f, x-a, y-b \in I$ и I не является главным идеалом, то $c = 0$. $\square$

Следствие 3.4. Пусть $X \subset \mathbf{A}^2$ — неприводимое аффинное многообразие, отличное от точки и всей плоскости $\mathbf{A}^2$. Тогда X — гиперповерхность, т.е. $I(X) = (f)$, где $f \in F[x,y]$ — неприводимый многочлен.

Теорема 3.4. Пусть $X \subseteq \mathbf{A}^2$ — аффинное многообразие. Тогда существует разложение $X = X_1 \cup \ldots \cup X_m$ в объединение конечного числа различных, не содержащихся друг в друге неприводимых аффинных многообразий. Это представление единственно с точностью до порядка.

Доказательство. Сначала покажем существование такого разложения. Возможны следующие случаи:

1. X — конечное множество. Тогда:

$$X = \bigcup_{i=1}^{m} \{P_i\}, \quad P_i = (a_i, b_i) \in \mathbf{A}^2, \quad i = 1, \ldots, m.$$

В этом случае:

$$X = \bigcup_{i=1}^{m} Z(x - a_i,\ y - b_i).$$

2. $X = \mathbf{A}^2$. В этом случае воспользуемся следствием 3.2.

3. X — бесконечное множество, $X \neq \mathbf{A}^2$. Пусть $f \in I(X)$, $f = f_1^{k_1} \ldots f_s^{k_s}$ — каноническое разложение многочлена f. Так как

$$(f) = \left(f_1^{k_1}\right) \ldots \left(f_s^{k_s}\right),$$

$$Z(f) = \bigcup_{i=1}^{s} Z(f_i^{k_i}) = \bigcup_{i=1}^{s} Z(f_i)$$

и $(f) \subseteq I(X)$, то по предложению 3.1

$$X \subseteq Z(f) = \bigcup_{i=1}^{s} Z(f_i).$$

Поэтому:

$$X = \bigcup_{i=1}^{s} (X \cap Z(f_i)) .$$

Пусть $X \cap Z(f_i)$ является бесконечным множеством при $i = 1, \ldots, r$ и конечным при $i = r+1, \ldots, s$. Зафиксируем произвольный многочлен $g \in I(X)$. Так как $X \subseteq Z(g)$, то $Z(g) \cap Z(f_i)$ является бесконечным множеством при $i = 1, \ldots, r$, а значит система уравнений $g(x, y) = f_i(x, y) = 0$ имеет бесконечное число решений. Поэтому из теоремы 3.2 следует, что f_i является делителем g. Следовательно, $(g) \subseteq (f_i)$, $Z(f_i) \subseteq Z(g)$. Так как:

$$X = \bigcap_{g \in I(X)} Z(g),$$

то $Z(f_i) \subseteq X$. Таким образом,

$$X = \left(\bigcup_{i=1}^{r} Z(f_i) \right) \cup \left(\bigcup_{i=r+1}^{s} (X \cap Z(f_i)) \right),$$

где второе объединение — конечное множество точек.

Покажем единственность разложения. Пусть существует два разложения:

$$X = X_1 \cup \ldots \cup X_m = Y_1 \cup \ldots \cup Y_n.$$

Тогда $X_1 \subseteq Y_1 \cup \ldots \cup Y_n$, поэтому $X_1 = (X_1 \cap Y_1) \cup \ldots \cup (X_1 \cap Y_n)$, где $X_1 \cap Y_i$ — алгебраические множества, $i = 1, \ldots, n$. Так как X_1 неприводимо, то для некоторого s выполнено равенство $X_1 = X_1 \cap Y_s$. Аналогично, $Y_s = X_t \cap Y_s$ для некоторого t. Получаем, что $X_1 \subseteq Y_s \subseteq X_t$. Из этого следует равенства $t = 1$ и $X_1 = Y_s$. И так далее. □

Пример 3.3. $Z(y^2 - x^2) = Z(y - x) \cup Z(y + x)$.

3.2. Проективная плоскость

Проективная прямая. Введем на множестве $F^2 \setminus (0, 0)$ отношение эквивалентности:

$$(X_1, Y_1) \sim (X_2, Y_2) \Leftrightarrow \exists \alpha \in F^* : X_2 = \alpha X_1,\ Y_2 = \alpha Y_1,$$

где $(X_1, Y_1), (X_2, Y_2) \in F^2 \setminus (0,0)$. Таким образом, класс эквивалентности, порожденный элементом $(X, Y) \in F^2 \setminus (0,0)$, имеет следующий вид: $\overline{(X,Y)} = \{(\alpha X, \alpha Y) \mid \alpha \in F^*\}$. Класс $\overline{(X,Y)}$ называется *проективной точкой.*

Множество точек аффинной плоскости

$$\left[\begin{array}{l} x = \alpha X, \\ y = \alpha Y, \ \alpha \in F^*, \end{array}\right.$$

представляет собой прямую $y = \dfrac{Y}{X}x$ при $X \neq 0$ либо прямую $x = 0$ при $X = 0$, с выколотой точкой $(0,0)$. Итак, класс $\overline{(X,Y)}$ это прямая в $\mathbf{A}^2(F)$, проходящая через точки $(0,0)$ и (X,Y), с выколотой точкой $(0,0)$. *Проективной прямой* $\mathbf{P}^1(F)$ над полем F называется множество классов $\overline{(X,Y)}$, $(X,Y) \in F^2 \setminus (0,0)$. Между множеством всех классов $\overline{(X,Y)} \in \mathbf{P}^1(F)$, у которых $Y \neq 0$, и множеством точек аффинной прямой $\mathbf{A}^1(F)$ можно установить взаимно однозначное соответствие:

$$\varphi\left(\overline{(X,Y)}\right) = \frac{X}{Y}, \quad Y \neq 0,$$

$$\varphi^{-1}(x) = \overline{(x,1)}, \quad x \in \mathbf{A}^1(F).$$

На проективной прямой еще имеется класс $\overline{(1,0)}$. Этот класс (проективная точка) называется *бесконечно удаленной точкой* и обозначается P_∞. С точностью до изоморфизма получаем такое равенство:

$$\mathbf{P}^1(F) = \mathbf{A}^1(F) \cup \{P_\infty\}.$$

Проективная плоскость. Аналогичным образом, на множестве $F^3 \setminus (0,0,0)$ введем отношение эквивалентности следующим образом:

$$(X_1, Y_1, Z_1) \sim (X_2, Y_2, Z_2) \Leftrightarrow \exists \alpha \in F^* :$$

$$X_2 = \alpha X_1, \ Y_2 = \alpha Y_1, \ Z_2 = \alpha Z_1.$$

Тогда $\overline{(X,Y,Z)} = \{(\alpha X, \alpha Y, \alpha Z) \mid \alpha \in F^*\}$ и каждый такой класс (проективная точка) представляет собой прямую в аффинном пространстве $\mathbf{A}^3(F)$, проходящую через точки $(0,0,0)$

и (X, Y, Z), с выколотой точкой $(0, 0, 0)$. *Проективной плоскостью* $\mathbf{P}^2(F)$ над полем F называется множество классов (проективных точек) $\overline{(X, Y, Z)}$, $(X, Y, Z) \in F^3 \setminus (0, 0, 0)$. Рассмотрим класс $\overline{(X, Y, Z)}$:

$$\left[\begin{array}{l} x = \alpha X, \\ y = \alpha Y, \quad \alpha \in F^*, \\ z = \alpha Z. \end{array}\right.$$

Если $Z \neq 0$, то пересечением класса $\overline{(X, Y, Z)}$ и плоскости $z = 1$ в аффинном пространстве $\mathbf{A}^3(F)$ является точка $\left(\dfrac{X}{Z}, \dfrac{Y}{Z}, 1\right)$. Поэтому отображение:

$$\varphi\left(\overline{(X, Y, Z)}\right) = \left(\frac{X}{Z}, \frac{Y}{Z}\right), \quad Z \neq 0,$$

$$\varphi^{-1}(x, y) = \overline{(x, y, 1)},$$

является биективным между аффинной плоскостью и всеми классами (проективными точками) проективной плоскости вида $\overline{(X, Y, 1)}$. При этом на проективной плоскости имеются точки вида $\overline{(X, Y, 0)}$. Эти точки называются бесконечно удаленными. С точностью до изоморфизма выполнено такое равенство:

$$\mathbf{P}^2(F) = \mathbf{A}^2(F) \cup \mathbf{P}^1(F).$$

Введем на проективной плоскости следующие операции:

$$\alpha\overline{(X, Y, Z)} = \overline{(\alpha X, \alpha Y, \alpha Z)},$$

$$\overline{(X_1, Y_1, Z_1)} + \overline{(X_2, Y_2, Z_2)} = \overline{(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2)},$$

где $\overline{(X, Y, Z)}, \overline{(X_1, Y_1, Z_1)}, \overline{(X_2, Y_2, Z_2)} \in \mathbf{P}^2(F)$, $\alpha \in F^*$. Нетрудно видеть, что данные операции определены корректно.

Каждой аффинной алгебраической кривой соответствует алгебраическая кривая на проективной плоскости следующим образом. Если аффинная алгебраическая кривая задана полиномом степени n вида $\sum_{0 \leqslant i+j \leqslant n} a_{ij} x^i y^j = 0$, то соответствующая проективная кривая задается полиномом степени n вида:

$$\sum_{0 \leqslant i+j \leqslant n} a_{ij} X^i Y^j Z^{n-i-j} = 0,$$

которому удовлетворяют соответствующие проективные точки: нужно заменить x на X/Z, y — на Y/Z и умножить на подходящую степень Z, чтобы освободиться от знаменателей. Этому уравнению удовлетворяют все проективные точки $\overline{(X,Y,Z)}$, $Z \neq 0$, для которых соответствующие аффинные точки (x,y), где $x = X/Z$, $y = Y/Z$, удовлетворяют уравнению:

$$\sum_{0 \leqslant i+j \leqslant n} a_{ij} x^i y^j = 0.$$

Помимо них проективной кривой могут удовлетворять бесконечно удаленные проективные точки.

Например, аффинной прямой $ax + by + c = 0$, $a^2 + b^2 > 0$, соответствует проективная прямая $aX + bY + cZ = 0$.

Пример 3.4. 1. Рассмотрим аффинную прямую $2x+y-1 = 0$. Ей соответствует проективная прямая $2X + Y - Z = 0$. Данная прямая состоит из всех проективных точек вида $\overline{(x, 1-2x, 1)}$, $x \in F$, и бесконечно удаленной точки $\overline{(1,-2,0)}$.

2. Аффинной кривой $y-x^2-1 = 0$ соответствует проективная кривая $YZ - X^2 - Z^2 = 0$, состоящая из точек проективных $\overline{(x, x^2+1, 1)}$, $x \in F$, и бесконечно удаленной точки $\overline{(0,1,0)}$.

3. Аффинной кривой $x^2 + y^2 = 1$ соответствует проективная кривая $X^2+Y^2 = Z^2$, состоящая из проективных точек $\overline{(x,y,1)}$, $x^2 + y^2 = 1$, $x, y \in F$. Ни одна бесконечно удаленная точка не принадлежит проективной кривой $X^2 + Y^2 = Z^2$.

Предложение 3.4. На проективной плоскости любые две несовпадающие прямые пересекаются ровно в одной точке.

Доказательство. Рассмотрим две параллельные аффинные прямые $ax+by+c = 0$ и $ax+by+d = 0$, $a^2+b^2 > 0$, $c \neq d$. Данным прямым соответствуют проективные прямые $aX + bY + cZ = 0$ и $aX + bY + dZ = 0$. Решением системы:

$$\begin{cases} aX + bY + cZ = 0 \\ aX + bY + dZ = 0 \end{cases}$$

при $a \neq 0$ будет точка $\overline{\left(-\frac{b}{a}, 1, 0\right)}$, а при $b \neq 0$ — точка

$\overline{\left(1, -\frac{a}{b}, 0\right)}$. При этом если $a \neq 0$ и $b \neq 0$, то выполнено равенство:

$$\overline{\left(-\frac{b}{a}, 1, 0\right)} = \overline{\left(1, -\frac{a}{b}, 0\right)}.$$

□

Замечание 3.2. Заметим, что любая проективная прямая $aX + bY + cZ = 0$, $a^2 + b^2 > 0$, содержит ровно одну бесконечно удаленную точку. Этой точкой является:
$\overline{(1, 0, 0)}$, если $a = 0$;
$\overline{(0, 1, 0)}$, если $b = 0$;
$\overline{(-b/a, 1, 0)} = \overline{(1, -a/b, 0)}$, если $a \neq 0$ и $b \neq 0$.

Пример 3.5 (**пифагоровы тройки**)**.** Хорошо известно, что для прямоугольного треугольника с катетами X, Y и гипотенузой Z выполнено равенство:

$$X^2 + Y^2 = Z^2. \tag{3.2}$$

Опишем все *пифагоровы тройки*, т.е. тройки целых чисел вида (X, Y, Z), для которых выполнено условие (3.2).

Обозначим через M множество всех пифагоровых троек:

$$M = \{(X, Y, Z) \in \mathbb{Z}^3 \mid X^2 + Y^2 = Z^2\}.$$

Определим на множестве M отношение эквивалентности следующим образом:

$$(X_1, Y_1, Z_1) \sim (X_2, Y_2, Z_2) \Leftrightarrow \exists q \in \mathbb{Q}^* :$$

$$X_2 = qX_1,\ Y_2 = qY_1,\ Z_2 = qZ_1.$$

Заметим, что если (X, Y, Z) — пифагорова тройка, то для любого целого m тройка чисел (mX, mY, mZ) также является пифагоровой.

Отметим также такое свойство элементов (X, Y, Z) множества M : *X, Y, Z взаимно просты тогда и только тогда, когда X, Y, Z попарно взаимно просты.* Действительно, пусть X, Y,

Z взаимно просты, но не попарно взаимно просты. Тогда какие-то два из чисел X, Y, Z делятся на некоторое простое число p. В силу равенства (3.2) третье число будет также делиться на p. Поэтому X, Y, Z не взаимно просты. Противоречие.

Исходя из сказанного, множество M разбивается на следующие непересекающиеся классы эквивалентных элементов:

$$M = \overline{(0,0,0)} \bigcup_{\substack{(X,Y,Z)\in M,\\ (X,Y)=(X,Z)=(Y,Z)=1}} \overline{(X,Y,Z)},$$

где $\overline{(X,Y,Z)} = \{(x,y,z) \in M \mid (X,Y,Z) \sim (x,y,z)\}$. Поэтому для описания всех решений уравнения (3.2) достаточно описать все элементы следующего множества:

$$M_1 = \{(X,Y,Z) \in M \mid (X,Y) = (X,Z) = (Y,Z) = 1\}.$$

Так как $(X,Y,0) \in M$ тогда и только тогда, когда $X = Y = 0$, то будем считать, что $Z \neq 0$. Преобразуем уравнение (3.2) к виду:

$$x^2 + y^2 = 1, \tag{3.3}$$

где $x = \dfrac{X}{Z}$, $y = \dfrac{Y}{Z}$. Данное уравнение задает окружность радиуса 1 с центром в начале координат (рис. 3.1).

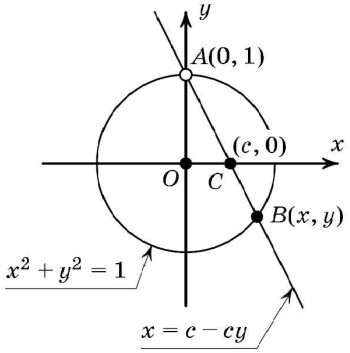


Рис. 3.1: Соответствие точек окружности точкам прямой $y = 0$

Рассмотрим следующие множества:

$$M_2 = \left\{ \left(\frac{a}{b}, \frac{c}{d}\right) \in \mathbb{Q}^2 \mid (a,b) = (c,d) = 1,\ \frac{a^2}{b^2} + \frac{c^2}{d^2} = 1 \right\},$$

$$M_3 = \left\{ \left(\frac{a}{b}, \frac{c}{b}\right) \in \mathbb{Q}^2 \mid (a,b) = (c,b) = 1,\ \frac{a^2}{b^2} + \frac{c^2}{b^2} = 1 \right\}.$$

Заметим, что $M_2 = M_3$. Действительно, пусть $\left(\frac{a}{b}, \frac{c}{d}\right) \in M_2$:

$$a^2d^2 + b^2c^2 = b^2d^2.$$

Так как $a^2d^2 = b^2(d^2 - c^2)$, то $b^2|a^2d^2$. При этом $(a^2, b^2) = 1$. Поэтому $b^2|d^2$. С другой стороны, $b^2c^2 = d^2(b^2 - a^2)$, поэтому $d^2|b^2c^2$. Так как $(c^2, d^2) = 1$, то $d^2|b^2$. Поэтому $M_2 = M_3$.

Заметим, что отображение $f : M_1 \to M_3$, определенное правилом $f(X, Y, Z) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$, является взаимно однозначным. Поэтому, с учетом равенства $M_2 = M_3$, исходная задача свелась к следующей: перечислить все рациональные точки (точки с рациональными координатами) полученной окружности. Зафиксируем некоторую рациональную точку окружности, например, $A = (0, 1)$. Проведем через точку A всевозможные прямые (кроме горизонтальной). Каждая такая прямая l пересечет окружность еще в одной точке $B = (x, y)$ и ось абсцисс в некоторой точке $C = (c, 0)$. Сопоставляя точке B точку C, получим взаимно однозначное соответствие между точками окружности (кроме A) и точками прямой $y = 0$. При таком соответствии точке $B = (x, y)$ будет соответствовать точка $c = \frac{x}{1 - y}$.

Заметим, что точка B имеет рациональные координаты тогда и только тогда, когда рационально число c. Действительно, из рациональности точки B рациональность c следует из отображения $c = \frac{x}{1 - y}$. Обратно, пусть число c рационально. Прямая, проходящая через точки A и C, определяется уравнением $x = c - cy$. Подставим его в уравнение окружности:

$$(c - cy)^2 + y^2 = 1,$$

$$(c^2 + 1)y^2 - 2c^2y + c^2 - 1 = 0.$$

Решая последнее квадратное уравнение, получаем, что либо $y = 1$ (что соответствует точке A), либо $y = \frac{c^2 - 1}{c^2 + 1}$, при этом

$x = c - cy = \dfrac{2c}{c^2+1}$. Поэтому если число c рациональна, то точка $B = \left(\dfrac{2c}{c^2+1}, \dfrac{c^2-1}{c^2+1}\right)$ также рациональна.

Таким образом, каждое рациональное решение уравнения (3.3), кроме $x = 0$, $y = 1$, получается, если в формулы:

$$x = \frac{2c}{c^2+1}, \quad y = \frac{c^2-1}{c^2+1}$$

подставить вместо c рациональное число. Подставим вместо c несократимую дробь $\dfrac{m}{n}$, m, n — целые числа. Тогда:

$$x = \frac{2c}{c^2+1} = \frac{2mn}{m^2+n^2}, \quad y = \frac{c^2-1}{c^2+1} = \frac{m^2-n^2}{m^2+n^2}, \quad m^2+n^2 > 0.$$

Возвращаемся к преобразованию $x = X/Z$, $y = Y/Z$:

$$\frac{X}{Z} = \frac{2mn}{m^2+n^2}, \quad \frac{Y}{Z} = \frac{m^2-n^2}{m^2+n^2}.$$

Дроби $x = X/Z$, $y = Y/Z$ несократимы, так как числа X, Y, Z попарно взаимно просты. Если бы дроби, стоящие в правых частях также были бы несократимы, то выполнялись бы равенства $X = 2mn$, $Y = m^2 - n^2$, $Z = m^2 + n^2$, но, например, при $m = 3$, $n = 1$ обе дроби сократимы. Нетрудно показать, что некоторое простое число p является делителем чисел $2mn$, $m^2 - n^2$, $m^2 + n^2$, $(m, n) = 1$, тогда и только тогда, когда $p = 2$, m и n — нечетные целые числа.

Таким образом, множество M_1 состоит из всех чисел вида:

$$X = mn, \quad Y = \frac{m^2-n^2}{2}, \quad Z = \frac{m^2+n^2}{2}$$

при взаимно простых нечетных m и n, $m > n > 0$, а также:

$$X = 2mn, \quad Y = m^2 - n^2, \quad Z = m^2 + n^2$$

при взаимно простых m и n, $m > n > 0$, одно из которых четно.

Пример 3.6 (**рациональные кривые**)**.** Решим также с помощью метода секущих следующую задачу: описать все тройки целых чисел X, Y, Z, для которых выполнено равенство:

$$X^2 + 2Y^2 = 3Z^2. \tag{3.4}$$

Заметим, что, как и ранее, $(X, Y, 0)$ является решением уравнения (3.4) тогда и только тогда, когда $Z = 0$. Поэтому решение $(0, 0, 0)$ далее рассматривать не будем. Разделим обе части равенства на Z^2 :

$$x^2 + 2y^2 = 3, \tag{3.5}$$

где $x = X/Z$, $y = Y/Z$. Данное уравнение описывает эллипс с полуосями $\sqrt{3}$ и $\dfrac{\sqrt{6}}{2}$ (рис. 3.2). Зафиксируем на данном эллипсе

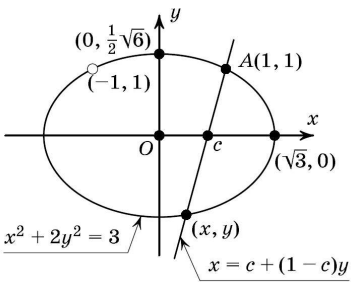


Рис. 3.2: Соответствие точек эллипса точкам прямой $y = 0$

произвольную рациональную точку, например, $A = (1, 1)$. Как и в предыдущем примере, установим взаимно однозначное соответствие между точками эллипса (кроме точки (-1,1)) и точками оси абсцисс. Уравнение прямой, проходящей через точки $A = (1, 1)$ и $(c, 0)$, можно записать в виде $x = (1 - c)y + c$. Подставим это в уравнение (3.5):

$$((1 - c)y + c)^2 + y^2 = 3,$$

$$((1 - c)^2 + 2)y^2 + 2c(1 - c)y + c^2 - 3 = 0.$$

Решая последнее квадратное уравнение относительно переменной y, получаем, что либо $y = 1$, либо:

$$y = \frac{c^2 - 3}{c^2 - 2c + 3}, \quad x = (1 - c)y + c = \frac{-c^2 + 6c - 3}{c^2 - 2c + 3}, \quad q \in \mathbb{Q}.$$

Пусть $c = m/n$. Тогда:

$$x = \frac{-m^2 + 6mn - 3n^2}{m^2 - 2mn + 3n^2}, \quad y = \frac{m^2 - 3n^2}{m^2 - 2mn + 3n^2}.$$

Целые решения уравнения (3.4) описываются формулами:

$$X = (-m^2+6mn-3n^2)q,\ Y = (m^2-3n^2)q,\ Z = (m^2-2mn+3n^2)q,$$

где m, n — целые числа, q — подходящее рациональное число.

3.3. Эллиптические кривые

Назовем (*плоской аффинной*) *кубической кривой* E_F (*кубикой*) над полем F алгебраическую кривую, заданную полиномом:

$$F(x,y) = \sum_{0\leqslant i+j\leqslant 3} \alpha_{ij} x^i y^j, \quad \alpha_{ij} \in F,$$

т.е. множество таких $(x,y) \in F^2$, для которых $F(x,y) = 0$. Соответствующая проективная кубика задается полиномом:

$$\sum_{0\leqslant i+j\leqslant 3} \alpha_{ij} X^i Y^j Z^{3-i-j}.$$

Определение 3.4. Точка $P = (x_0, y_0) \in E_F$ называется *простой (неособой)*, если значения частных производных:

$$\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial x}(x_0,y_0), \quad \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial y}(x_0,y_0)$$

в точке P не равны нулю одновременно. В противном случае точка называется *кратной (особой)*.

Определение 3.5. Кубика E_F называется *гладкой*, если все ее точки являются простыми.

Предложение 3.5. Пусть $char\, F \neq 2$ и $char\, F \neq 3$. Кубика E_F, заданная многочленом:

$$F(x,y) = y^2 - x^3 - ax - b,$$

является гладкой тогда и только тогда, когда многочлен $f(x) = x^3 + ax + b \in F[x]$ не имеет кратных корней в поле разложения.

Доказательство. Нам понадобятся следующие простые факты.

- Если $x_0 \in F$ — корень многочлена $f(x)$, то $(x_0, y) \in E_F$ тогда и только тогда, когда $y = 0$.
- $(x_0, 0) \in F^2$ принадлежит кубике E_F тогда и только тогда, когда x_0 — корень многочлена $f(x)$.
- x_0 — кратный корень многочлена $f(x)$ тогда и только тогда, когда $f'(x_0) = 0$.

Пусть x_0 — кратный корень многочлена $f(x)$. Так как:

$$\frac{\partial F}{\partial x} = -3x^2 - a = -f'(x), \quad \frac{\partial F}{\partial y} = 2y,$$

то:

$$\frac{\partial F}{\partial x}(x_0, 0) = -f'(x_0) = 0, \quad \frac{\partial F}{\partial y}(x_0, 0) = 0.$$

Поэтому кубика E_F не является гладкой.

Обратно, пусть кубика E_F не является гладкой. Тогда найдется точка $P = (x_0, y_0) \in E_F$, для которой:

$$\frac{\partial F}{\partial x}(P) = 0 = \frac{\partial F}{\partial y}(P).$$

Так как $\frac{\partial F}{\partial y} = 2y$, то $y_0 = 0$. Поэтому x_0 — корень многочлена $f(x)$, причем:

$$\frac{\partial F}{\partial x}(x_0, 0) = -f'(x_0) = 0.$$

Поэтому x_0 — кратный корень. □

Предложение 3.6. Пусть $char\, F \neq 2$, $char\, F \neq 3$, $a, b \in F$. Многочлен $f(x) = x^3 + ax + b$ не имеет кратных корней в поле разложения тогда и только тогда, когда $4a^3 + 27b^2 \neq 0$.

Доказательство. Напомним, что любой многочлен $g(x)$ не имеет кратных корней в поле разложения тогда и только тогда, когда $(g(x), g'(x)) = 1$ (см. [13]). Применим для нахождения $(f(x), f'(x))$ алгоритм Евклида:

$$(f(x), f'(x)) = (x^3 + ax + b,\ 3x^2 + a) = (3x^2 + a,\ \frac{2}{3}ax + b). \quad (3.6)$$

Рассмотрим следующие случаи:

1. $a = 0$. Тогда $(f(x), f'(x)) = (3x^2, b)$. В этом случае $(f(x), f'(x)) = 1$ тогда и только тогда, когда $b \neq 0$, а это, в свою очередь, равносильно условию $4a^3 + 27b^2 \neq 0$.

2. $b = 0$. Учитывая (3.6), получаем:

$$(f(x), f'(x)) = (3x^2 + a, \frac{2}{3}ax).$$

В этом случае $(f(x), f'(x)) = 1$ тогда и только тогда, когда $a \neq 0$, что, в свою очередь, равносильно условию $4a^3 + 27b^2 \neq 0$.

3. $a \neq 0$, $b \neq 0$. С учетом равенства:

$$3x^2 + a = \left(\frac{2}{3}ax + b\right)\left(\frac{9}{2a}x - \frac{27b}{4a^2}\right) + \frac{4a^3 + 27b^2}{4a^2},$$

получаем:

$$(f(x), f'(x)) = \left(\frac{2}{3}ax + b,\ \frac{4a^3 + 27b^2}{4a^2}\right).$$

Поэтому $(f(x), f'(x)) = 1$ тогда и только тогда, когда $4a^3 + 27b^2 \neq 0$. □

Следствие 3.5. Кубика E_F, заданная многочленом $F(x, y) = y^2 - x^3 - ax - b$, является гладкой тогда и только тогда, когда $4a^3 + 27b^2 \neq 0$.

Определение 3.6. Пусть характеристика поля F отлична от 2 и 3 и кубический многочлен $x^3 + ax + b$, $a, b \in F$, не имеет кратных корней (равносильно условию $4a^3 + 27b^2 \neq 0$). *Эллиптической кривой* над полем F называется множество всех таких точек $(x, y) \in F^2$, удовлетворяющие уравнению:

$$y^2 = x^3 + ax + b. \tag{3.7}$$

Если характеристика поля F равна 2, то эллиптической кривой называется множество точек, удовлетворяющих уравнению либо типа:

$$y^2 + cy = x^3 + ax + b,$$

либо типа:

$$y^2 + xy = x^3 + ax^2 + b,$$

где кубические многочлены в правых частях могут иметь кратные корни. Если же характеристика поля F равна 3, то эллиптической кривой называется множество точек, удовлетворяющих уравнению типа:

$$y^2 = x^3 + ax^2 + bx + c, \tag{3.8}$$

где кубический многочлен в правой части не имеет кратных корней.

Пусть $F = \mathbb{R}$. Рассмотрим кубику E, заданную уравнением вида:

$$y^2 = x^3 + ax + b, \quad a, b \in \mathbb{R}.$$

Так как $y = \pm\sqrt{x^3 + ax + b}$, то график кубики E симметричен относительно оси абсцисс. Для нахождения точек пересечения данной кубики с осью абсцисс необходимо решить уравнение:

$$x^3 + ax + b = 0. \tag{3.9}$$

Это можно сделать с помощью формул Кардано. Дискриминант этого уравнения имеет вид:

$$D = \left(\frac{a}{3}\right)^3 + \left(\frac{b}{2}\right)^2.$$

При $D < 0$ уравнение (3.9) имеет три различных действительных корня α, β и γ; если $D = 0$, то (3.9) имеет три действительных корня, один и которых имеет кратность 2; если $D > 0$, то (3.9) имеет один действительный корень и два комплексно сопряженных. Все три случая приведены на рис. 3.3. При этом случай $D = 0$ не подходит для задания эллиптической кривой.

Замечание 3.3. Имеется общая форма Вейерштрасса уравнения эллиптической кривой, которая применима при любом поле:

$$E: \quad y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6; \tag{3.10}$$

в случае $char\ F \neq 2$ ее можно привести к виду $y^2 = x^3 + ax^2 + bx + c$ (или к виду $y^2 = x^3 + bx + c$, если $char\ F > 3$). В случае $char\ F = 2$ данное уравнение можно привести к виду $y^2 + cy = x^3 + ax + b$ либо к виду $y^2 + xy = x^3 + ax^2 + b$.

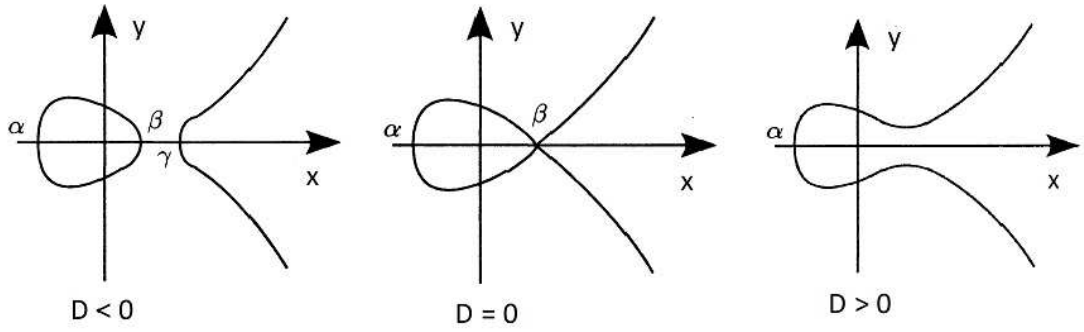


Рис. 3.3: Вид эллиптических кривых при D<0, D=0, D>0

Замечание 3.4. Эллиптическая кривая должна быть неособой в том смысле, что частные производные $\partial E/\partial x$, $\partial E/\partial y$ функции:

$$E(x,y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$$

не должны обращаться в ноль одновременно ни в одной точке, удовлетворяющих уравнению $E(x,y) = 0$.

В предложении 3.5 показано, что условие отсутствия кратных корней у кубического многочлена в правой части в (3.7) эквивалентно требованию, чтобы все точки кривой были неособыми.

Нетрудно также показать, что условие отсутствия кратных корней у кубического многочлена в правой части в (3.8) эквивалентно требованию, чтобы все точки кривой были неособыми.

Замечание 3.5. Для эллиптической кривой, заданной уравнением (3.10), вводятся следующие константы:

$$\begin{aligned}
b_2 &= a_1^2 + 4a_2,\\
b_4 &= a_1a_3 + 2a_4,\\
b_6 &= a_3^2 + 4a_6,\\
b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2,\\
c_4 &= b_2^2 - 24b_4,\\
c_6 &= -b_2^3 + 36b_2b_4 - 216b_6.
\end{aligned}$$

Дискриминант кривой E определяется по формуле:

$$D = D(E) = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6.$$

Если $char\ F \neq 2$ и $char\ F \neq 3$, то $D(E) = -16(4a^3 + 27b^2)$.

В предложении 3.6 показано, что кривая E неособа тогда и только тогда, когда $D \neq 0$.

Сложение точек эллиптической кривой над полем $\mathbb{R}$. Проективная эллиптическая кривая E с условием $D \neq 0$ будет иметь такой вид:

$$Y^2Z = X^3 + aXZ^2 + bZ^3.$$

Помимо проективных точек вида $\overline{(x, y, 1)}$, удовлетворяющих уравнению $y^2 = x^3 + ax + b$, данной кривой также будет принадлежать ровно одна бесконечно удаленная проективная точка $\mathcal{O} = \overline{(0, 1, 0)}$.

Пусть P и Q точки эллиптической кривой E. Определим операции $-P$ и $P + Q$.

1. Если $P = \mathcal{O}$, то $-P = \mathcal{O}$, $P + Q = Q$. Далее полагаем, что $P \neq \mathcal{O}$, $Q \neq \mathcal{O}$.

2. Для $P = (x, y)$ точка $-P = (x, -y)$. Понятно, что $-P \in E$.

3. Пусть $P = (x_1, y_1)$, $Q = (x_2, y_2)$.

3.1. Если $x_1 \neq x_2$, то прямая, проходящая через точки P и Q имеет ровно три точки пересечения с кривой E, две из которых это P и Q. Обозначим третью точку пересечения через $A = (x, y)$. Пусть $B = -A$ (данная операция определена в предыдущем пункте). Тогда полагаем $P + Q = B$.

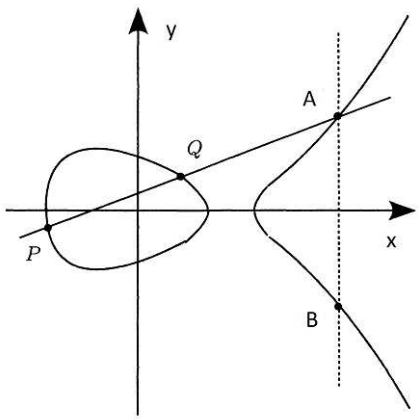


Рис. 3.4: Сложение точек B=P+Q

3.2. Если $P = -Q$, то проективная прямая, проходящая через точки P и Q, имеет вид $X = x_1Z$ и пересекает проективную кривую E в точках P, Q и $\mathcal{O}$. Поэтому полагаем $P + Q = \mathcal{O}$.

3.3. Если $P = Q$ (будем обозначать $P + P = [2]P$), то рассмотрим касательную l к кривой E в точке P.

3.3.1. Если l параллельна оси ординат ($y_1 = 0$), то полагаем $P + P = \mathcal{O}$.

3.3.2. Если $y_1 \neq 0$, то касательная l пересекает кривую E ровно в двух точках: P и A. Тогда полагаем $P + P = -A$.

Теперь покажем, что точка пересечения кривой E и прямой l, проходящей через точки $P = (x_1, y_1)$ и $Q = (x_2, y_2)$, действительно существует и отлична от P и Q. Пусть $x_1 \neq x_2$. Тогда уравнение прямой l имеет такой вид:

$$y = \frac{y_2 - y_1}{x_2 - x_1}x + y_1 - x_1\frac{y_2 - y_1}{x_2 - x_1}.$$

Пусть $k = \dfrac{y_2 - y_1}{x_2 - x_1}$. Перепишем уравнение прямой l в виде $y = y_1 + k(x - x_1)$. Подставив это уравнение в (3.7), получаем:

$$(y_1 + k(x - x_1))^2 = x^3 + ax + b.$$

В итоге, получается такой вид уравнения:

$$x^3 - k^2x^2 + \ldots = 0.$$

Известно, что сумма коней нормированного кубического уравнения равна коэффициенту при x^2, взятому с обратным знаком (теорема Виета для кубических уравнений), поэтому:

$$x_1 + x_2 + x_3 = k^2,$$

откуда:

$$x_3 = k^2 - x_1 - x_2, \quad y_3 = y_1 + k(x_3 - x_1).$$

В итоге:

$$P + Q = (k^2 - x_1 - x_2,\ k(x_1 - x_3) - y_1).$$

Пусть теперь $P = Q$. Продифференцируем (3.7) по x :

$$2yy' = 3x^2 + a.$$

Угловой коэффициент равен значению производной в точке P :

$$\widetilde{k} = \frac{3x_1^2 + a}{2y_1}.$$

Поэтому уравнение касательной l к графику кривой E в точке P примет вид $y = y_1 + \widetilde{k}(x - x_1)$. Аналогичные рассуждения показывают, что:

$$P + P = [2]P = (\widetilde{k}^2 - 2x_1,\ \widetilde{k}(x_1 - x_3) - y_1).$$

Пример 3.7. Рассмотрим эллиптическую кривую $E:\ y^2 = x^3 - 4x + 1$. Для точек $P = (-2, 1), Q = (-1, 2) \in E$ найдем $P + Q$ и $[2]P$:

$$P + Q = (4, -7),\quad [2]P = (20, -89).$$

Введенная бинарная операция «+» превращает множество точек эллиптической кривой (с учетом точки $\mathcal{O}$) в аддитивную абелеву группу. Введем следующие обозначения для любого целого числа n :

$$[n]P = \underbrace{P + \ldots + P}_{n},$$
$$[0]P = \mathcal{O},$$
$$[-n]P = \underbrace{-P - \ldots - P}_{n}.$$

Сложение точек эллиптической кривой над произвольным полем характеристики, большей трех. Введенные выше формулы для $P + Q$ и $[2]P$ останутся также верными, если в качестве поля F будет выступать любое поле с характеристикой, отличной от 2 и 3. Например, если $F = \mathbb{Z}_p$, $p > 3$, то эллиптическая кривая $E_p(a, b)$ примет такой вид:

$$y^2 = x^3 + ax + b \pmod{p},$$

где $a, b \in \mathbb{Z}_p$ и:

$$4a^3 + 27b^2 \not\equiv 0 \pmod{p}.$$

Пример 3.8. Пусть $F = \mathbb{Z}_7$. Рассмотрим эллиптическую кривую $E_7(3, 1) : \ y^2 = x^3 + 3x + 1 \pmod 7$. Для точек $P = (5, 1), Q = (6, 2) \in E$ найдем $P + Q$ и $[2]P$:

$$P + Q = (4, 0),\quad [2]P = (6, 2).$$

Эллиптические кривые над конечным полем. Пусть $GF(q)$ — конечное поле из q элементов, $q = p^n$, E — эллиптическая кривая, определенная над $GF(q)$. Нетрудно видеть, что эллиптическая кривая может иметь не более $2q + 1$ различных $GF(q)$-точек, т.е. точку $\mathcal{O}$ и не более $2q$ точек (x, y), $x, y \in GF(q)$, удовлетворяющих уравнению эллиптической кривой (которое зависит от характеристики поля — см. определение 3.6). Но так как лишь у половины элементов поля $GF(q)$ имеются квадратные корни, то количество таких точек примерно вдвое меньше числа $2q + 1$.

Вычисление количества точек эллиптической кривой не всегда возможно. Известна асимптотически точная формула количества точек эллиптической кривой над конечным полем.

Теорема 3.5 (Хассе). Пусть N — число $GF(q)$-точек эллиптической кривой, определенной над $GF(q)$. Тогда:

$$|N - (q + 1)| \leqslant 2\sqrt{q}.$$

Существуют эллиптические кривые E, точки которой образуют циклическую группу с порождающим элементом $G \in E$, т.е.:

$$E = \{G,\ [2]G, \ldots, [n]G = \mathcal{O}\},$$

где $n = |E|$. Число точек кривой E также может быть простым числом. В этом случае любая точка $G \in E \setminus \{\mathcal{O}\}$ является порождающим элементом циклической группы $(E, +)$.

Пример 3.9. Пусть $F = \mathbb{Z}_5$. Рассмотрим эллиптическую кривую $E_5(1, 3)$: $y^2 = x^3 + x + 3 \pmod 5$. Найдем все точки этой кривой с координатами из F. Заметим, что для данной кривой $D = 4$. Для удобства вычисления квадратных корней построим такую таблицу:

$$\begin{array}{lccccc} x \in F: & 0 & 1 & 2 & 3 & 4 \\ x^2 \in F: & 0 & 1 & 4 & 4 & 1 \end{array}$$

Теперь будем подставлять элементы поля F вместо переменной x и, по возможности, вычислять квадратные корни: $y^2 =$

$0^3 + 0 + 3 = 3$. В нижней строке предыдущей таблицы вычета 3 нет, значит на эллиптической кривой $E_5(1,3)$ нет точки с абсциссой, равной 0. Дальше $y^2 = 1^3 + 1 + 3 = 0$. Учитывая предыдущую таблицу, получаем, что $(1,0) \in E_5(1,3)$. Продолжая процесс дальше, получаем, что $(4,1), (4,4) \in E_5(1,3)$. Итак:

$$E_5(1,3) = \{\mathcal{O},\ (1,0),\ (4,1),\ (4,4)\}.$$

Данное множество с введенной выше операцией сложения является аддитивной абелевой группой. Результаты сложения несложно получить по полученным выше формулам. Сведем их в таблицу:

$+$	$\mathcal{O}$	$(1,0)$	$(4,1)$	$(4,4)$
$\mathcal{O}$	$\mathcal{O}$	$(1,0)$	$(4,1)$	$(4,4)$
$(1,0)$	$(1,0)$	$\mathcal{O}$	$(4,4)$	$(4,1)$
$(4,1)$	$(4,1)$	$(4,4)$	$(1,0)$	$\mathcal{O}$
$(4,4)$	$(4,4)$	$(4,1)$	$\mathcal{O}$	$(1,0)$

Используя данную таблицу, найдем порядок каждой точки эллиптической кривой $E_5(1,3)$.

$$[2]P(1,0) = (1,0) + (1,0) = \mathcal{O}.$$

Порядок точки $(1,0)$ равен двум.

$$\begin{aligned}
[2]P(4,1) &= (4,1) + (4,1) = (1,0),\\
[3]P(4,1) &= (1,0) + (4,1) = (4,4),\\
[4]P(4,1) &= (4,4) + (4,1) = \mathcal{O}.
\end{aligned}$$

Порядок точки $(4,1)$ равен четырем.

$$\begin{aligned}
[2]P(4,4) &= (4,4) + (4,4) = (1,0),\\
[3]P(4,4) &= (1,0) + (4,4) = (4,1),\\
[4]P(4,4) &= (4,1) + (4,4) = \mathcal{O}.
\end{aligned}$$

Порядок точки $(4,4)$ равен четырем.

Из данных вычислений видно, что $E_5(1,3)$ — конечная циклическая группа порядка 4, а в роли образующего элемента можно выбрать либо точку $(4,1)$, либо точку $(4,4)$.

Пример 3.10. Как и в предыдущем примере, найдем все точки эллиптической кривой $E_7(2,3):\ y^2 = x^3 + 2x + 3 \pmod 7$ на полем $F = \mathbb{Z}_7$. Используя прием предыдущего примера, получаем:

$$E_7(2,3) = \{\mathcal{O},\ (2,1),\ (2,6),\ (3,1),\ (3,6),\ (6,0)\}.$$

В следующей таблице представлены результаты сложения точек кривой $E_7(2,3)$:

$+$	$\mathcal{O}$	$(2,1)$	$(2,6)$	$(3,1)$	$(3,6)$	$(6,0)$
$\mathcal{O}$	$\mathcal{O}$	$(2,1)$	$(2,6)$	$(3,1)$	$(3,6)$	$(6,0)$
$(2,1)$	$(2,1)$	$(3,6)$	$\mathcal{O}$	$(2,6)$	$(6,0)$	$(3,1)$
$(2,6)$	$(2,6)$	$\mathcal{O}$	$(3,1)$	$(6,0)$	$(2,1)$	$(3,6)$
$(3,1)$	$(3,1)$	$(2,6)$	$(6,0)$	$(3,6)$	$\mathcal{O}$	$(2,1)$
$(3,6)$	$(3,6)$	$(6,0)$	$(2,1)$	$\mathcal{O}$	$(3,1)$	$(2,6)$
$(6,0)$	$(6,0)$	$(3,1)$	$(3,6)$	$(2,1)$	$(2,6)$	$\mathcal{O}$

Каждая точка эллиптической кривой $E_7(2,3)$ имеет соответствующие порядки:

$$(2,1) - \ 6, \quad (2,6) - \ 6, \quad (3,1) - \ 3, \quad (3,6) - \ 3, \quad (6,0) - \ 2.$$

$E_7(2,3)$ — конечная циклическая группа порядка 6, а в роли образующего элемента можно выбрать либо точку $(2,1)$, либо точку $(2,6)$.

Аналогом возведения в степень $a^n \pmod p$ является вычисление значения:

$$[n]P = \underbrace{P + \ldots + P}_{n},$$

причем существует быстрый алгоритм вычисления данного выражения, аналогичный быстрому алгоритму возведения в степень. Пусть:

$$n = n_t 2^t + \ldots + n_1 2 + n_0, \quad n_i \in \{0,1\},\ i = 0, \ldots, t.$$

Тогда:

$$[n]P = n_0 P + n_1 [2]P + \ldots + n_t [2^t]P,$$

где:

$$n_i[2^i]P = \begin{cases} \mathcal{O}, & n_i = 0, \\ [2^i]P, & n_i = 1. \end{cases}$$

При этом:

$$[2^{i-1}]P + [2^{i-1}]P = [2^i]P.$$

Поэтому количество операций сложения вы вычислении значения $[n]P$ сверху ограничено числом $2 \cdot \log_2 n$.

Пусть известны точки эллиптической кривой P и $Q = [n]P$. Задача отыскания числа n называется задачей дискретного логарифмирования на эллиптической кривой.

3.4. Сложение точек эллиптической кривой над произвольным полем

Рассмотрим общий вид эллиптической кривой (3.10). Проективная эллиптическая кривая E будет иметь такой вид:

$$E: \quad Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

Секущая, проходящая через бесконечно удаленную точку $\mathcal{O} = \overline{(0,1,0)}$ и данную точку кривой E, является вертикальной прямой.

Пусть P и Q — точки эллиптической кривой E. Определим операции $-P$ и $P + Q$.

1. Если $P = \mathcal{O}$, то $-P = \mathcal{O}$, $P + Q = Q$. Далее полагаем, что $P \neq \mathcal{O}$, $Q \neq \mathcal{O}$.

2. Для $P = (x, y)$ точка $-P = (x, -y - a_1x - a_3)$. Действительно, пусть $A = (x, y_1)$, $B = (x, y_2) \in E$. Тогда:

$$y_1^2 + a_1xy_1 + a_3y_1 = y_2^2 + a_1xy_2 + a_3y_2,$$

$$(y_1 - y_2)(y_1 + y_2) + a_1x(y_1 - y_2) + a_3(y_1 - y_2) = 0,$$

$$(y_1 - y_2)(y_1 + y_2 + a_1x + a_3) = 0,$$

$$y_2 = -y_1 - a_1x - a_3.$$

3. Пусть $P = (x_1, y_1)$, $Q = (x_2, y_2)$.

3.1. Если $x_1 \neq x_2$, то прямая, проходящая через точки P и Q имеет ровно три точки пересечения с кривой E, две из которых это P и Q. Уравнение прямой l имеет такой вид:

$$y = \frac{y_2 - y_1}{x_2 - x_1}x + y_1 - x_1\frac{y_2 - y_1}{x_2 - x_1}.$$

Пусть $k = \dfrac{y_2 - y_1}{x_2 - x_1}$. Перепишем уравнение прямой l в виде $y = y_1 + k(x - x_1)$. Подставив это уравнение в (3.10), получаем:

$$(y_1 + k(x - x_1))^2 + a_1x(y_1 + k(x - x_1)) + a_3(y_1 + k(x - x_1)) =$$

$$= x^3 + a_2x^2 + a_4x + a_6.$$

В итоге, получается такой вид уравнения:

$$x^3 + (a_2 - k^2 - a_1k)x^2 + \ldots = 0.$$

Известно, что сумма коней нормированного кубического уравнения равна коэффициенту при x^2, взятому с обратным знаком (теорема Виета для кубических уравнений), поэтому:

$$x_1 + x_2 + x_3 = k^2 + a_1k - a_2,$$

откуда:

$$x_3 = k^2 + a_1k - a_2 - x_1 - x_2, \quad y_3 = -(y_1 + k(x_3 - x_1)) - a_1x_3 - a_3.$$

В итоге:

$$P + Q = (k^2 + a_1k - a_2 - x_1 - x_2, -(k + a_1)x_3 - y_1 + kx_1 - a_3).$$

3.2. Если $P = -Q$, то проективная прямая, проходящая через точки P и Q, имеет вид $X = x_1Z$ и пересекает проективную кривую E в точках P, Q и $\mathcal{O}$. Поэтому полагаем $P + Q = \mathcal{O}$.

3.3. Если $P = Q$, то рассмотрим касательную l к кривой E в точке P.

3.3.1. Если $2y_1 + a_1x_1 + a_3 = 0$, то полагаем $P + P = \mathcal{O}$.

3.3.2. Если $2y_1 + a_1x_1 + a_3 \neq 0$, то касательная l пересекает кривую E ровно в двух точках. Продифференцируем (3.10) по x :

$$2yy' + a_1y + a_1xy' + a_3y' = 3x^2 + 2a_2x + a_4.$$

Угловой коэффициент равен значению производной в точке P :

$$\widetilde{k} = \frac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3}.$$

Поэтому уравнение касательной l к графику кривой E в точке P примет вид $y = y_1 + \widetilde{k}(x - x_1)$. Аналогичные рассуждения показывают, что:

$$P+P = [2]P = (\widetilde{k}^2+a_1\widetilde{k}-a_2-x_1-x_2, -(\widetilde{k}+a_1)x_3-y_1+\widetilde{k}x_1-a_3).$$

Глава 4. Математические модели открытых текстов

4.1. Детерминированная модель

Пусть $A = \{a_1, a_2, \ldots, a_n\}$ — некоторое конечное множество, которое будем называть алфавитом. Открытый текст представляет собой конечную последовательность $x_1 x_2 \ldots x_l$ символов алфавита A. Такую последовательность будем также называть *словом.* Число l называется длиной открытого текста.

Элемент алфавита A называется *буквой*, элемент множества A^2 (т.е. слово длины 2) называется *биграммой*, элемент множества A^3 — триграммой и т.д. В целом, элемент множества A^m называется *m-граммой.*

Обозначим через A^* множество всех конечных слов в алфавите A. Также обозначим через Λ пустое слово. Будем считать, что $\Lambda \in A^*$. Для слов x и y из множества A^* определим их произведение как слово xy, получающееся приписыванием слова y справа к слову x. Например, если $x = x_1 \ldots x_m$, $y = y_1 \ldots y_n$, где $x_1, \ldots, x_m, y_1, \ldots, y_n \in A$, то:

$$xy = x_1 \ldots x_m y_1 \ldots y_n.$$

При этом для любого $x \in A^*$ верны такие равенства:

$$x\Lambda = \Lambda x = x.$$

Множество A^* с введенной операцией умножения будет являться моноидом.

Пусть S — некоторое фиксированное подмножество в A^*. Определим два подмножества D_0 и D_1 в множестве A^* следующим образом:

$$D_0 = \{xsy \mid x, y \in A^*,\ s \in S\},$$

$$D_1 = A^* \backslash D_0.$$

Так как $\Lambda \in A^*$, то $S \subseteq D_0$. Заметим также, что для любых $x, y \in A^*$ и любого $d_0 \in D_0$ слово xd_0y принадлежит множеству D_0. В частности, множество D_0 замкнуто относительно операции умножения, чего нельзя сказать о множестве D_1.

Множество D_0 можно определить также следующим образом: множество D_0 состоит из всех таких элементов $x \in A^*$, для каждого из которых найдется такой элемент $s \in S$, что s является подсловом слова x.

Детерминированная модель открытых текстов определяется разбиением множества A^* на два непересекающихся подмножества D_0 и D_1, которые соответственно называются множеством запретных и множеством разрешенных последовательностей. При этом последовательность $x_1 \dots x_l \in A^*$ является открытым текстом в детерминированной модели тогда и только тогда, когда $x_1 \dots x_l \in D_1$.

Заметим, что конструкция множества D_0 позволяет дать такой эквивалентный критерий проверки открытого текста: последовательность $x_1 \dots x_l \in A^*$ является открытым текстом тогда и только тогда, когда ни один из элементов множества S не является подсловом в $x_1 \dots x_l$.

Множество S называется *множеством запретных m-грамм*, $m = 2, 3, \dots$ Данное множество порождает множество D_0. Множество S удобнее определять таким образом, чтобы никакое слово меньшей длины не являлось подсловом слова большей длины, т.е. для любых слов $s_1 \dots s_k$, $\widetilde{s}_1 \dots \widetilde{s}_l \in S$, $k < l$, слово $s_1 \dots s_k$ не является подсловом в $\widetilde{s}_1 \dots \widetilde{s}_l$.

Пример 4.1. Пусть A — русский алфавит. В русских текстах биграмма ьь является запретной, т.е. ьь $\in S$. Также запретными являются биграммы ъъ, оь, аъ и т.д. Определив таким образом множество запретных m-грамм S, тем самым определяем детерминированную модель открытых текстов в русском языке.

4.2. Вероятностная модель

В такой модели открытый текст рассматривается как случайная последовательность символов алфавита A. Вероятность случайного события $a_{i_1} \dots a_{i_l} \in A^*$ определяется как вероятность такой последовательности событий:

$$P(a_{i_1} \dots a_{i_l}) = P(x_1 = a_{i_1},\ \dots, x_l = a_{i_l}).$$

Зафиксируем произвольное значение $l \in \mathbb{N}$. Тройка $(\Omega = A^l,\ F,\ P)$, где Ω — пространство элементарных событий, F — алгебра событий, P — вероятностная мера на F, образует вероятностное пространство, описывающее эксперимент по появлению слова $x_1 \dots x_l$, причем должны быть выполнены следующие условия:

1. Для любого $x_1 \dots x_l \in A^l$:

$$0 \leqslant P(x_1 \dots x_l) \leqslant 1;$$

2.

$$\sum_{x_1 \dots x_l \in A^l} P(x_1 \dots x_l) = 1.$$

При этом l-мерные распределения вероятностей множества A^l $(A^l,\ F,\ P)$ должны удовлетворять следующему условию согласования.

3. Для любого $k > l$:

$$P(x_1 \dots x_l) = \sum_{x_{l+1} \dots x_k \in A^{k-l}} P(x_1 \dots x_l x_{l+1} \dots x_k),$$

т.е. вероятность слова $x_1 \dots x_l$ длины l есть сумма вероятностей всех продолжений этого слова до длины k.

Таким образом, задавая распределение вероятностей на множествах A^l, $l \geqslant 1$, с условием согласования, тем самым определяется вероятностная модель открытых текстов.

Рассмотрим основные вероятностные модели.

4.2.1. Вероятностная модель независимых символов алфавита

Такая модель предполагает, что открытым текстом является последовательность независимых букв алфавита A. В такой модели вероятность открытого текста $x_1 \dots x_l$ определяется таким равенством:

$$P(x_1 \dots x_l) = \prod_{i=1}^{l} P(x_i), \tag{4.1}$$

где для любого $a \in A$ $P(a) > 0$ и $\sum_{a \in A} P(a) = 1$.

Понятно, что с таким определением вероятности будут выполнены условия 1—3 предыдущего пункта.

Из формулы (4.1) видно, что распределения вероятностей на множествах A^l, $l \geqslant 2$, полностью определяются распределением вероятностей на множестве A. Вероятности $P(a_1),\dots, P(a_n)$ из распределения $P(A)$ вычисляются априорно на некотором исходном материале.

Например, ниже приведены две таблицы, в которых указаны примерные частоты букв русского и английского языков (символом «-» обозначен символ-разделитель слов в открытых текстах):

Таблица 1

Частоты букв в русском языке

-	О	Е	А	И	Т	Н	С
0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
Р	В	Л	К	М	Д	П	У
0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
Я	Ы	З	Ь,Ъ	Б	Г	Ч	Й
0,018	0,016	0,016	0,014	0,014	0,012	0,012	0,010
Х	Ж	Ю	Ш	Ц	Щ	Э	Ф
0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

Таблица 2

Частоты букв в английском языке

-	E	T	A	O	N	I	S	R
0,185	0,097	0,076	0,064	0,062	0,057	0,056	0,052	0,047
H	L	D	C	U	P	F	M	W
0,040	0,031	0,028	0,025	0,018	0,018	0,018	0,017	0,016
Y	B	G	V	K	Q	X	J	Z
0,015	0,013	0,013	0,007	0,004	0,002	0,002	0,001	0,001

К недостатку такой модели можно отнести то обстоятельство, что любая конечная последовательность символов в алфавите A будет являться открытым текстом с ненулевой вероятностью, что не учитывает особенности языков, содержащие запретные m-граммы.

4.2.2. Вероятностная модель независимых биграмм

В такой модели любой открытый текст рассматривается как последовательность независимых биграмм из множества A^2. Вероятность открытого текста $x_1x_2\dots x_{2l-1}x_{2l} \in A^{2l}$ определяется таким равенством:

$$P(x_1x_2\dots x_{2l-1}x_{2l}) = \prod_{i=1}^{l} P(x_{2i-1}x_{2i}), \tag{4.2}$$

где для любой биграммы $ab \in A^2$ $P(ab) \geqslant 0$ и $\sum_{ab\in A^2} P(ab) = 1$.

С таким определением вероятности данная модель также удовлетворяет условиям 1—3.

Распределения вероятностей на множествах A^{2l}, $l \geqslant 2$, полностью определяются распределением вероятностей на множестве A^2. Вероятности $P(a_ia_j)$, $1 \leqslant i, j \leqslant n$, из распределения $P(A^2)$ также вычисляются априорно на некотором исходном текстовом материале. При этом результаты записываются в такую таблицу:

	a_1	a_2	$\dots$	a_n
a_1	$P(a_1a_1)$	$P(a_1a_2)$	$\dots$	$P(a_1a_n)$
a_2	$P(a_2a_1)$	$P(a_2a_2)$	$\dots$	$P(a_2a_n)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
a_n	$P(a_na_1)$	$P(a_na_2)$	$\dots$	$P(a_na_n)$

После этого данная таблица используется для вычисления вероятностей произвольной последовательности в алфавите A четной длины с помощью формулы (4.2).

Заметим, что данная модель точнее, по сравнению с предыдущей, отражает особенности естественных языков. Теперь если первая буква запретной биграммы располагается на нечетной позиции в сообщении, то данное сообщение будет иметь нулевую вероятность. В то же время если первая буква запретной биграммы будет располагаться на четной позиции в сообщении, то данная биграмма будет проигнорирована.

4.2.3. Вероятностная модель марковски зависимых букв

Пусть имеется некоторый исходный материал открытых текстов. На данном исходном материале строятся две таблицы распределения вероятностей на множестве A :

A	a_1	a_2	$\ldots$	a_n
P	$P_A(a_1)$	$P_A(a_2)$	$\ldots$	$P_A(a_n)$

и на множестве A^2 :

	a_1	a_2	$\ldots$	a_n
a_1	$P_{A^2}(a_1a_1)$	$P_{A^2}(a_1a_2)$	$\ldots$	$P_{A^2}(a_1a_n)$
a_2	$P_{A^2}(a_2a_1)$	$P_{A^2}(a_2a_2)$	$\ldots$	$P_{A^2}(a_2a_n)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
a_n	$P_{A^2}(a_na_1)$	$P_{A^2}(a_na_2)$	$\ldots$	$P_{A^2}(a_na_n)$

Причем для любых $a, b \in A$ $P_A(a) > 0$ и $P_{A^2}(ab) \geqslant 0$. Также должны выполняться равенства:

$$\sum_{a \in A} P_A(a) = 1, \quad \sum_{ab \in A^2} P_{A^2}(ab) = 1.$$

Данная модель открытых текстов представляет собой *однородную цепь Маркова.* Вероятность сообщения $x_1 \ldots x_l \in A^l$ в данной модели определяется по следующей формуле:

$$P(x_1 \ldots x_l) = P_A(x_1) \prod_{i=2}^{l} P(x_i \mid x_{i-1}),$$

где условная вероятность $P(b \mid a)$ вычисляется таким образом:

$$P(b \mid a) = \frac{P_{A^2}(ab)}{P_A(a)}.$$

Данная модель еще более точно отражает особенности естественных языков по сравнению с моделью независимых биграмм. Например, если сообщение содержит запретную биграмму, то вероятность данного сообщения будет равна нулю.

Глава 5. Шифры замены и перестановки (исторические шифры)

5.1. Одноалфавитные шифры замены

5.1.1. Шифр простой замены

Пусть имеются некоторые конечные алфавиты A и B, в которых записываются соответственно открытые и шифрованные тексты, причем $|A| = |B|$. Установим некоторым образом взаимно однозначное соответствие φ из A в B ($\varphi : A \to B$). Пусть $x_1x_2 \dots x_l$ — некоторый открытый текст, где все $x_i \in A$. Тогда процесс зашифрования сообщения $x_1x_2 \dots x_l$ состоит в замене каждой буквы x_i на соответствующую букву $\varphi(x_i)$, т.е. после зашифрования сообщения $x_1x_2 \dots x_l$ получим шифрованный текст

$$y_1y_2 \dots y_l = \varphi(x_1)\varphi(x_2) \dots \varphi(x_l).$$

Обратно, если $y_1y_2 \dots y_l$ — некоторый шифрованный текст, где все $y_i \in B$, то расшифрование данного шифртекста выглядит следующим образом:

$$x_1x_2 \dots x_l = \varphi^{-1}(y_1)\varphi^{-1}(y_2) \dots \varphi^{-1}(y_l),$$

где φ^{-1} — взаимно однозначное отображение из B в A, обратное к отображению φ.

Ключом шифра будет являться отображение φ.

Пусть $A = B$. В этом случае шифр простой замены можно представить как числовые преобразования символов открытого текста следующим образом. Сопоставим каждой букве алфавита $A = \{a_0, a_1, \dots, a_{n-1}\}$ ее порядковый номер:

$$\begin{array}{cccc} a_0 & a_1 & \dots & a_{n-1} \\ 0 & 1 & \dots & n-1. \end{array}$$

Пусть φ — некоторая подстановка полученного множества A :

$$\begin{pmatrix} 0 & 1 & \dots & n-1 \\ \varphi(0) & \varphi(1) & \dots & \varphi(n-1) \end{pmatrix}.$$

Обозначим:

$$d_i = \varphi(i) - i \pmod{n},\ i = 0, 1, \dots, n-1.$$

Тогда процесс преобразования сообщения $x_1 x_2 \dots x_l$, где все $x_i \in \{0, 1, \dots, n-1\}$, в шифрованное сообщение $y_1 y_2 \dots y_l$ можно представить следующим образом:

$$y_i = x_i + d_{x_i} \pmod{n},\ i = 1, 2, \dots, l.$$

5.1.2. Шифр сдвига

Шифр сдвига является частным случаем шифра простой замены. В данном случае $A = B$, а в качестве подстановок φ выступают циклические подстановки. Понятно, что число таких подстановок равно мощности алфавита A.

Как и в предыдущем пункте, сопоставим каждой букве алфавита A ее порядковый номер. Зафиксируем некоторое значение $d \in \{0, 1, \dots, n-1\}$. Тогда процесс преобразования сообщения $x_1 x_2 \dots x_l$, где все $x_i \in \{0, 1, \dots, n-1\}$, в шифрованное сообщение $y_1 y_2 \dots y_l$ можно представить следующим образом:

$$y_i = x_i + d \pmod{n},\ i = 1, 2, \dots, l.$$

Очевидно, что процесс расшифрования можно представить таким образом:

$$x_i = y_i + n - d \pmod{n},\ i = 1, 2, \dots, l.$$

Покажем, как можно взломать шифр сдвига, опираясь на статистические характеристики того или иного языка. Предположим, что мы перехватили некоторое шифрованное сообщение и нам известна следующая информация: данное сообщение зашифровано с помощью шифра сдвига и известно априорное распределение символов алфавита A (см., например, таблицы 1.1 и 1.2):

$$\begin{array}{ll} A: & 0 \quad 1 \quad \dots \quad n-1 \\ P: & p_0 \quad p_1 \quad \dots \quad p_{n-1} \end{array}$$

И пусть в данной таблице символ i имеет наибольшую вероятность p_i. Пусть $y = y_1 \dots y_m$ — перехваченное шифрованное сообщение. Подсчитаем в нем число вхождений каждого символа из алфавита A :

$$\begin{array}{cccc} 0 & 1 & \dots & n-1 \\ f_0 & f_1 & \dots & f_{n-1}, \end{array}$$

где $f_0 + f_1 + \ldots + f_{n-1} = m$. Выберем из данной таблицы такое значение $j \in A$, которое имеет наибольшее значение f_j. Тогда закономерно предположить, что символ i при шифровании с помощью ключа d преобразовался в символ $j : j = i+d \pmod n$. Поэтому очень вероятно, что $d = j - i \pmod n$.

Заметим, что шифр простой замены также легко вскрывается при помощи частотного анализа. Для этого нужно сопоставить частоты появлений букв шифртекста с частотами появлений букв того языка, к которому относится открытый текст. После чего наиболее часто встречаемые буквы шифртекста заменяются на наиболее часто встречаемые буквы алфавита, а остальные замены происходят исходя из вероятности появления того или иного слова и знания синтаксических правил используемого языка.

5.1.3. Улучшенный криптоанализ шифра сдвига

Нам понадобится следующее утверждение.

Предложение 5.1. Пусть $x_1, \dots, x_n$ — произвольный набор из n действительных чисел, где n — любое натуральное число. Тогда:

(i) для любой перестановки $\alpha \in S_n$, где S_n — симметрическая группа степени n, выполнено неравенство:

$$\sum_{i=1}^{n} x_i x_{\alpha(i)} \leqslant \sum_{i=1}^{n} x_i^2;$$

(ii) если перестановка $\alpha \in S_n$ обладает тем свойством, что найдется такой элемент $i_0 \in \{1, 2, \dots, n\}$, что $x_{\alpha(i_0)} \neq x_{i_0}$, то

будет выполнено строгое неравенство:

$$\sum_{i=1}^{n} x_i x_{\alpha(i)} < \sum_{i=1}^{n} x_i^2; \tag{5.1}$$

(iii) если все элементы $x_1, \ldots, x_n$ попарно различны, то для любой нетождественной перестановки $\alpha \in S_n \backslash \{e\}$ будет выполнено строгое неравенство (5.1).

Доказательство. (i) Для начала заметим, что для любых $x, y \in \mathbb{R}$ выполнено такое неравенство:

$$xy \leqslant \frac{x^2 + y^2}{2}, \tag{5.2}$$

которое следует из такого выражения:

$$x^2 + y^2 - 2xy = (x - y)^2 \geqslant 0.$$

Также для любой перестановки $\alpha \in S_n$ отметим такое очевидное равенство:

$$\sum_{i=1}^{n} x_{\alpha(i)}^2 = \sum_{i=1}^{n} x_i^2. \tag{5.3}$$

Применяя (5.2) и (5.3), получаем доказательство пункта (i):

$$\sum_{i=1}^{n} x_i x_{\alpha(i)} \overset{(5.2)}{\leqslant} \frac{1}{2} \sum_{i=1}^{n} (x_i^2 + x_{\alpha(i)}^2) = \frac{1}{2} \left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_{\alpha(i)}^2 \right) \overset{(5.3)}{=}$$

$$\overset{(5.3)}{=} \frac{1}{2} \left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i^2 \right) = \sum_{i=1}^{n} x_i^2.$$

(ii) Пусть $x, y \in \mathbb{R}$ и $x \neq y$. Тогда:

$$x^2 + y^2 - 2xy = (x - y)^2 > 0.$$

Поэтому для таких элементов x и y выполнено следующее строгое неравенство:

$$xy < \frac{x^2 + y^2}{2}. \tag{5.4}$$

Пусть некоторая перестановка $\alpha \in S_n$ обладает тем свойством, что найдется такой элемент $i_0 \in \{1, 2, \ldots, n\}$, что $x_{\alpha(i_0)} \neq x_{i_0}$.

Тогда:

$$\sum_{i=1}^{n} x_i x_{\alpha(i)} = x_{i_0} x_{\alpha(i_0)} + \sum_{\substack{1 \leqslant i \leqslant n, \\ i \neq i_0}} x_i x_{\alpha(i)} \overset{(5.2),\ (5.4)}{<}$$

$$< \frac{x_{i_0}^2 + x_{\alpha(i_0)}^2}{2} + \frac{1}{2} \sum_{\substack{1 \leqslant i \leqslant n, \\ i \neq i_0}} (x_i^2 + x_{\alpha(i)}^2) =$$

$$= \frac{x_{i_0}^2}{2} + \frac{1}{2} \sum_{\substack{1 \leqslant i \leqslant n, \\ i \neq i_0}} x_i^2 + \frac{x_{\alpha(i_0)}^2}{2} + \frac{1}{2} \sum_{\substack{1 \leqslant i \leqslant n, \\ i \neq i_0}} x_{\alpha(i)}^2 =$$

$$= \frac{1}{2} \sum_{i=1}^{n} x_i^2 + \frac{1}{2} \sum_{i=1}^{n} x_{\alpha(i)}^2 \overset{(5.3)}{=} \sum_{i=1}^{n} x_i^2.$$

(iii) Доказательство данного пункта следует из пункта (ii): для любой нетождественной перестановки $\alpha \in S_n \backslash \{e\}$ обязательно найдется такой элемент $i_0 \in \{1, 2, \ldots, n\}$, зависящий от α, что $\alpha(i_0) \neq i_0$, а так как элементы $x_1, \ldots, x_n$ попарно различны, то $x_{\alpha(i_0)} \neq x_{i_0}$. □

Пусть имеется некоторый источник открытых сообщений, который вырабатывает открытые тексты в алфавите $A = \mathbb{Z}_n$, причем известны частоты появления символов из A :

$$\begin{array}{lcccc} A: & 0 & 1 & \ldots & n-1 \\ P: & p_0 & p_1 & \ldots & p_{n-1}. \end{array}$$

Предположим, что открытые тексты, вырабатываемые данным источником сообщений, шифруются шифром сдвига.

Пусть $y = y_1 \ldots y_m$ — перехваченное шифрованное сообщение. Подсчитаем в нем число вхождений каждого символа из алфавита A :

$$\begin{array}{cccc} 0 & 1 & \ldots & n-1 \\ f_0 & f_1 & \ldots & f_{n-1}, \end{array}$$

где $f_0 + f_1 + \ldots + f_{n-1} = m$.

Пусть $S_n \ni T^j$ — циклическая перестановка на j позиций влево. Рассмотрим суммы вида:

$$\sum_{i=0}^{n-1} p_i \cdot f_{T^j(i)} = \sum_{i=0}^{n-1} p_i \cdot f_{i+n-j \pmod n}, \quad j = 0, 1, \ldots, n-1.$$

Из всех полученных сумм выберем сумму с наибольшим значением. Пусть эта сумма имеет значение $j = d$. Из предложения 5.1 следует, что очень вероятно, что d и есть ключ зашифрования, причем:

$$\sum_{i=0}^{n-1} p_i \cdot \frac{f_{T^d(i)}}{m} \approx \sum_{i=0}^{n-1} p_i^2.$$

5.1.4. Аффинный шифр

Усовершенствуем шифр сдвига с помощью аффинных отображений:

$$y = ax + b \pmod n,$$

где a и b — некоторые фиксированные целые числа из кольца вычетов $\mathbb{Z}_n$, причем $(a, n) = 1$.

Пусть $x_1 \ldots x_l$ — некоторый открытый текст, где все $x_i \in \mathbb{Z}_n$. Тогда из данного сообщения получается шифртекст $y_1 \ldots y_l$ следующим образом:

$$y_i = ax_i + b \pmod n, \ i = 1, 2, \ldots, l.$$

Так как $(a, n) = 1$, то отображение $f : \mathbb{Z}_n \to \mathbb{Z}_n$, определяемое формулой:

$$y = f(x) = ax + b \pmod n, \ x \in \mathbb{Z}_n,$$

является взаимно однозначным. Обратным к отображению f будет отображение f^{-1} :

$$x = f^{-1}(y) = \widetilde{a}y + \widetilde{b} \pmod n, \ y \in \mathbb{Z}_n,$$

где $\widetilde{a} = a^{-1} \in \mathbb{Z}_n$, $b = -a^{-1}b$. Поэтому расшифрование сообщения $y_1 \ldots y_l$ выглядит следующим образом:

$$x_i = (y_i + n - b)a^{-1} \pmod n, \ i = 1, 2, \ldots, l.$$

Заметим, что шифр сдвига является частным случаем аффинного шифра при $a = 1$.

Пусть нам известно, что перехваченное сообщение зашифровано с помощью аффинного шифра букв n-буквенного алфавита. Наша цель — определить значение ключа a, b. Применим частотный анализ. Пусть в открытых текстах буквы t_1 и t_2 имеют наибольшие вероятности появления. Находим в шифртексте две наиболее часто встречаемые буквы. Обозначим их через s_1 и s_2. Получаем такую систему:

$$\begin{cases} t_1 a + b = s_1 \pmod n \\ t_2 a + b = s_2 \pmod n. \end{cases}$$

Решив данную систему относительно a и b, найдем значение искомого ключа шифрования.

Также можно провести и улучшенный криптоанализ аффинного шифра на основе предложения 5.1. Пусть $a_1, \ldots, a_k$ — все обратимые элементы в $\mathbb{Z}_n$ (заметим, что $k = \varphi(n)$, где φ — функция Эйлера). Рассмотрим все суммы вида:

$$\sum_{i=0}^{n-1} p_i \cdot f_{a_s \cdot (i+n-j) \pmod n}, \quad s = 1, \ldots, k, \quad j = 0, \ldots, n-1. \quad (5.5)$$

Выберем такие значения a_s и j, при которых сумма вида (5.5) имеет максимальное значение. Тогда из предложения 5.1 следует, что наиболее вероятно значения $a = (a_s)^{-1}$ и $b = j$ и будут являться значениями ключа аффинного шифра.

5.1.5. Преобразование биграмм аффинным шифром

Для начала заметим, что для любых натуральных n и k отображение:

$$f : \mathbb{Z}_n^k \to \mathbb{Z}_{n^k},$$

определенное по правилу:

$$f(x_1, \ldots, x_k) = x_1 n^{k-1} + x_2 n^{k-2} + \ldots + x_k, \quad (x_1, \ldots, x_k) \in \mathbb{Z}_n^k,$$

является биекцией. Рассмотрим случай $k = 2$:

$$f : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_{n^2}, \quad f(u, v) = un + v, \quad (u, v) \in \mathbb{Z}_n \times \mathbb{Z}_n.$$

Пусть $A = \mathbb{Z}_n$, $B = \mathbb{Z}_{n^2}$. Разобьем открытый текст x на двухбуквенные блоки, называемые *биграммами.* Если сообщение состоит из нечетного числа букв, то добавим к нему еще одну букву, причем с тем условием, чтобы не исказить смысл сообщения. Поэтому пусть:

$$x = x_1 x_2 \dots x_{2l-1} x_{2l} = X_1 \dots X_l, \quad X_i = x_{2i-1} x_{2i}.$$

Пусть $a, b \in \mathbb{Z}_{n^2}$ и $(a, n^2) = 1$. Тогда аффинное преобразование $y = ax + b$ множества $\mathbb{Z}_{n^2}$ является обратимым. Зашифруем открытый текст x по правилу:

$$y_i = a \cdot f(X_i) + b \pmod{n^2}, \quad i = 1, \dots, l.$$

Тогда расшифрование будет проходить следующим образом:

$$X_i = f^{-1}(a^{-1}(y_i + n^2 - b) \pmod{n^2}).$$

Аналогично можно рассматривать случаи, когда $k = 3, 4, \dots$

5.2. Многоалфавитные шифры замены

5.2.1. Шифр замены с конечным ключом

В рассмотренных предыдущих случаях применяется только один алфавит шифртекста, поэтому данные шифры относятся к моноалфавитным шифрам, которые, как мы видели, легко взламываются. Существуют шифры, в которых используется целый набор алфавитов шифртекста. Такие шифры называются многоалфавитными и позволяют в некоторой степени скрыть естественную частоту появления букв в тексте.

Пусть все открытые тексты записываются в конечном алфавите A. Пусть также имеются конечные множества B_1, B_2,..., B_s и биективные преобразования:

$$\varphi_i : A \to B_i, \ i = 1, 2, \dots, s.$$

Тогда открытый текст $x_1 \dots x_s x_{s+1} \dots x_{2s} \dots$ в процессе шифрования преобразуется в шифртекст:

$$\varphi_1(x_1) \dots \varphi_s(x_s) \varphi_1(x_{s+1}) \dots \varphi_s(x_{2s}) \dots \ .$$

Пусть $A = \{a_1, \ldots, a_n\}$, $B_i = \{b_1^i, \ldots, b_n^i\}$, $i = 1, \ldots, s$. Зафиксируем перестановки $\sigma_1, \ldots, \sigma_s \in S_n$, где S_n — симметрическая группа порядка n. Тогда биективные преобразования φ_i, $i = 1, \ldots, s$, можно представить в виде подстановок:

$$\varphi_i : \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ b^i_{\sigma_i(1)} & b^i_{\sigma_i(2)} & \ldots & b^i_{\sigma_i(n)} \end{pmatrix}$$

Частным случаем является ситуация, когда:

$$A = B_1 = \ldots = B_s.$$

В этом случае преобразования $\varphi_i : A \to A$, $i = 1, \ldots, s$, можно представить в виде подстановок:

$$\varphi_i : \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ a_{\sigma_i(1)} & a_{\sigma_i(2)} & \ldots & a_{\sigma_i(n)} \end{pmatrix},$$

где $\sigma_1, \ldots, \sigma_s \in S_n$.

5.2.2. Шифр Виженера

Данный шифр является частным случаем шифра замены с конечным ключом. Пусть A — некоторый алфавит, в котором записываются открытые и шифрованные тексты. Как и ранее, сопоставив каждой букве алфавита A ее порядковый номер, можно считать, что $A = \{0, 1, \ldots, n-1\}$. Для зашифрования открытого текста используется ключ — слово (в общем случае, хаотичный набор букв) в алфавите A. Чтобы зашифровать некоторое сообщение, поступаем следующим образом. Пусть $d_1 \ldots d_s$ — ключевое слово. Под каждой буквой открытого текста помещается буква ключа. Ключ повторяется необходимое количество раз:

$$\begin{array}{cccccccc} x_1 & \ldots & x_s & x_{s+1} & \ldots & x_{2s} & \ldots \\ d_1 & \ldots & d_s & d_1 & \ldots & d_s & \ldots \end{array}$$

Тогда шифртекст $y_1 \ldots y_s y_{s+1} \ldots y_{2s} \ldots$ получается следующим образом:

$$y_1 = x_1 + d_1 \pmod n, \ldots, \; y_s = x_s + d_s \pmod n,$$

$$y_{s+1} = x_{s+1} + d_1 \pmod n, \ldots$$

Если буквы открытого текста и буквы ключевого слова занумеровать, начиная с нуля ($x_0 \dots x_{l-1}$, $d_0 \dots d_{s-1}$), то шифр Виженера можно описать следующим образом:

$$y_i = x_i + d_{i \pmod s} \pmod n, \quad i = 0, 1, \dots, l-1.$$

5.2.3. Криптоанализ шифра Виженера

Несмотря на то, что многоалфавитные шифры маскируют естественную частоту появления букв в шифрованном тексте и данные шифры являются более стойкими, нежели моноалфавитные шифры, все же частотный анализ применим и здесь. Предположим, что мы знаем длину ключа (ниже приведены два метода определения длины ключа: метод Казиски и по индексу совпадения). Пусть она равна s. Разобьем перехваченный шифртекст на блоки одинаковой длины в s символов. Тогда каждая первая буква полученных блоков шифровалась на одной и той же букве ключевого слова. То же самое верно и для остальных букв в блоках, имеющих одинаковые номера. Поэтому следует применить частотный анализ (например, улучшенный криптоанализ шифра сдвига) для каждой из s полученных групп в отдельности.

Метод Казиски. В открытом тексте встречаются одинаковые сочетания символов (биграммы, триграммы и т.д.). При шифровании может так получиться, что эти одинаковые сочетания зашифрованы одинаковой частью ключевого слова. Поэтому и в шифртексте появятся одинаковые сочетания символов. При этом заметим, что случайно такие одинаковые сочетания могут появиться с достаточно малой вероятностью.

Пусть $d_1, d_2, \dots$ — расстояния между повторениями некоторой m-граммы в шифртексте. Вычислим наибольший общий делитель d данных чисел. Так как число s делит d_1, d_2,..., то по свойству наибольшего общего делителя оно делит и d. Чем больше повторений имеет данная m-грамма в шифртексте, тем более вероятно, что d совпадет с s. Если какая-то другая m-грамма повторяется в шифртексте с расстояниями $\rho_1, \rho_2, \dots$ и ρ

— наибольший общий делитель данных чисел, то число s будет также делить число (d, ρ).

Наиболее эффективно анализировать повторения триграмм в шифртексте, так как вероятность присутствия в шифртексте одинаковых m-грамм при $m > 3$ крайне мала, а при анализе повторений биграмм велик процент случайных совпадений.

После того, как выдвинута гипотеза о значении длины ключа s, ее можно проверить с помощью *индекса совпадения*, введенного в практику У. Фридманом в 1920 году.

Индекс совпадения. Пусть $x = x_1 \dots x_m$ — некоторое слово длины m в алфавите $A = \{a_1, \dots, a_n\}$. Индексом совпадения в x будем называть вероятность того, что две случайно выбранные буквы из x окажутся одинаковыми, и обозначать $I_c(x)$.

Теорема 5.1. Пусть $x = x_1 \dots x_m$ — некоторое слово длины m в алфавите $A = \{a_1, \dots, a_n\}$ и f_i — число вхождений буквы a_i в x, $i = 1, \dots, n$. Тогда индекс совпадения в слове x вычисляется по следующей формуле:

$$I_c(x) = \frac{1}{m(m-1)} \sum_{i=1}^{n} f_i(f_i - 1).$$

Доказательство. В качестве испытания в данном случае выступает случайный выбор двух букв в слове x. Поэтому пространством элементарных исходов будут все сочетания по две буквы из m-буквенного слова x :

$$\Omega = \{x_1x_2,\ x_1x_3, \dots, x_{m-1}x_m\}.$$

Общее число элементарных исходов равно, очевидно, C_m^2.

Пусть A_i — событие, которое заключается в том, что две случайно выбранные буквы из x окажутся одинаковыми и равны a_i $(i = 1, \dots, n)$. Поскольку в слове x ровно f_i букв a_i, то событию A_i будут благоприятствовать $C_{f_i}^2$ элементарных исходов. Поэтому вероятность события A_i равна следующему значению:

$$P(A_i) = \frac{C_{f_i}^2}{C_m^2} = \frac{f_i(f_i - 1)}{m(m-1)}.$$

Так как $I_c(x) = P(A_1 + \ldots + A_n)$ и события $A_1, \ldots, A_n$ являются несовместными, то по теореме сложения вероятностей имеем:

$$I_c(x) = P(A_1 + \ldots + A_n) = P(A_1) + \ldots + P(A_n) =$$

$$= \frac{f_1(f_1-1)}{m(m-1)} + \ldots + \frac{f_n(f_n-1)}{m(m-1)} = \frac{1}{m(m-1)} \sum_{i=1}^{n} f_i(f_i-1).$$

□

Заметим очень важный момент. Пусть φ — некоторая подстановка множества A. Применим к слову $x = x_1 \ldots x_m$ шифр простой замены:

$$y = y_1 \ldots y_m = \varphi(x_1) \ldots \varphi(x_m).$$

Тогда $I_c(x) = I_c(y)$. Действительно, пусть:

$$\begin{array}{cccc} a_1 & \ldots & a_n \\ f_1 & \ldots & f_n \end{array}$$

— частоты появлений букв алфавита A в слове x. Тогда для слова y эти частоты будут выглядеть следующим образом:

$$\begin{array}{cccc} \varphi(a_1) & \ldots & \varphi(a_n) \\ f_1 & \ldots & f_n \end{array}$$

Поэтому, принимая во внимание теорему 5.1, приходим к равенству $I_c(x) = I_c(y)$.

Рассмотрим модель открытых текстов со стационарным источником независимых букв алфавита. Пусть буквы $a_1, \ldots, a_n$ появляются в открытых текстах с соответствующими вероятностями $p_1, \ldots, p_n$. В такой модели открытого текста вероятность того, что две случайно выбранные буквы слова $x = x_1 \ldots x_m$ окажутся одинаковыми и равны a_i, будет равна p_i^2. Так как при достаточно большом значении m имеют место приближенные равенства:

$$\frac{f_i}{m} \approx p_i, \quad \frac{f_i - 1}{m-1} \approx p_i,$$

то:

$$I_c(x) \approx \sum_{i=1}^{n} p_i^2.$$

Например, для английского и русского языков индекс совпадения приближенно равен соответственно значениям 0,066 и 0,0529.

Пусть $y = y_1 \dots y_m$ — перехваченный шифртекст и s — предполагаемая длина ключевого слова. Если число s является истинной длиной ключа, то каждая из последовательностей символов:

$$Y_1 = y_1 y_{s+1} y_{2s+1} \dots,$$
$$Y_2 = y_2 y_{s+2} y_{2s+2} \dots,$$
$$\dots$$
$$Y_s = y_s y_{2s} y_{3s} \dots$$

шифровалась на одной букве ключа. Поэтому, с учетом сказанного выше, для любого $i = 1, \dots, s$ должно выполняться приближенное равенство:

$$I_c(Y_i) \approx \sum_{i=1}^{n} p_i^2.$$

Например, для англоязычного текста должно выполняться приближенное равенство $I_c(Y_i) \approx 0,066$ для любого $i = 1, \dots, s$.

Если же число s отлично от истинного значения длины ключа, то буквы алфавита A в Y_i для любого $i = 1, \dots, s$ будут иметь «более» равномерное распределение, так как в этом случае буквы из Y_i являются результатом шифрования многоалфавитного шифра замены. Поэтому $I_c(Y_i)$ будет ближе к числу $\frac{1}{n}$ (для англоязычного текста — к 0,038), так как если буквы в некотором слове z имеют равномерное распределение, то:

$$I_c(z) \approx \sum_{i=1}^{n} \frac{1}{n^2} = \frac{1}{n}.$$

Приведем небольшое обоснование этому факту. Пусть $x = x_1 \dots x_m$ — некоторое слово в алфавите $A = \mathbb{Z}_n$ и f_i — число вхождений буквы a_i в x, $i = 1, \dots, n$. Обозначим:

$$I(x) = \frac{1}{m^2} \sum_{i=1}^{n} f_i^2.$$

Понятно, что $I(x) \approx I_c(x)$ для больших m. Также обозначим через $\varphi_a : \mathbb{Z}_n \to \mathbb{Z}_n$ отображение, определенное следующим правилом: $\varphi_a(x) = a + x \pmod n$.

Предложение 5.2. Для любого слова $x = x_1 \dots x_m$ в алфавите $\mathbb{Z}_n$ и любых $a, b \in \mathbb{Z}_n$ выполнено неравенство

$$I(\varphi_a(x)\varphi_b(x)) \leqslant I(\varphi_a(x)\varphi_a(x)).$$

Доказательство. Подсчитаем количество вхождений букв алфавита $\mathbb{Z}_n$ соответственно в слова $\varphi_a(x)$ и $\varphi_b(x)$:

$$\begin{array}{cccccc} 0 & 1 & \dots & n-1 & \qquad & 0 & 1 & \dots & n-1 \\ f_0 & f_1 & \dots & f_{n-1} & & g_0 & g_1 & \dots & g_{n-1}, \end{array}$$

при этом последовательности $f_0, \dots, f_{n-1}$ и $g_0, \dots, g_{n-1}$ отличаются друг от друга циклическим сдвигом, поэтому:

$$\sum_{i=0}^{n-1} f_i^2 = \sum_{i=0}^{n-1} g_i^2.$$

Из ранее сказанного следует, что $I(\varphi_a(x)) = I(\varphi_b(x)) = I(x)$. Поэтому:

$$I(\varphi_a(x)\varphi_b(x)) = \frac{1}{4m^2}\sum_{i=0}^{n-1}(f_i + g_i)^2 = \frac{1}{4m^2}\sum_{i=0}^{n-1}(f_i^2 + 2f_ig_i + g_i^2) \leqslant$$

$$\overset{(5.2)}{\leqslant} \frac{1}{4m^2}\sum_{i=0}^{n-1}(2f_i^2 + 2g_i^2) = \frac{1}{4m^2}\sum_{i=0}^{n-1}(4f_i^2) =$$

$$= I(\varphi_a(x)\varphi_a(x)) = I(\varphi_a(x)). \qquad \square$$

Заметим, что если в предыдущем предложении найдется такое $i \in \{0, 1, \dots, n-1\}$, что $f_i \neq g_i$, то будет выполнено строгое неравенство $I(\varphi_a(x)\varphi_b(x)) < I(\varphi_a(x)\varphi_a(x))$.

Исходя из неравенства

$$(x_1 + \dots + x_k)^2 \leqslant k(x_1^2 + \dots + x_k^2),$$

которое выполнено для любых действительных чисел $x_1, \dots, x_k$, предложение 5.2 можно обобщить следующим образом.

Предложение 5.3. Для любого слова $x = x_1 \dots x_m$ в алфавите $\mathbb{Z}_n$ и любых $a_1, \dots, a_k \in \mathbb{Z}_n$ выполнено неравенство

$$I(\varphi_{a_1}(x) \dots \varphi_{a_k}(x)) \leqslant I(\underbrace{\varphi_{a_1}(x) \dots \varphi_{a_1}(x)}_{k}).$$

5.2.4. Многопетлевые подстановки

Если в шифре Виженера используется только один ключ, то в данном шифре — несколько ключей зашифрования. Их называют петлевыми или первичными ключами.

Многопетлевый шифр описывается следующей формулой:

$$y_i = x_i + d_i^1 \pmod{s_1} + d_i^2 \pmod{s_2} + \dots + d_i^k \pmod{s_k} \pmod{n},$$

$$i = 0, 1, \dots, l-1,$$

где $x_0 \dots x_{l-1}$ — открытый текст в алфавите $A = \mathbb{Z}_n$, $d_0^j \dots d_{s_j-1}^j$ — j-й первичный ключ длины s_j, $j = 1, \dots, k$.

Последовательное и циклическое применение первичных ключей дает, в итоге, составной ключ. Таким образом, многопетлевый шифр с первичными ключами — это шифр Виженера с составным ключом. Период составного ключа равен наименьшему общему кратному длин всех первичных ключей. Поэтому если длины первичных ключей являются попарно взаимно простыми числами, то длина составного ключа равна их произведению.

5.2.5. Аффинный блочный шифр

Рассмотрим аффинные преобразования $\mathbb{Z}_n$-модуля $\mathbb{Z}_n^m$, состоящего из векторов-столбцов $(x_1, \dots, x_m)^T$, где все $x_i \in \mathbb{Z}_n$:

$$Y = AX + B, \tag{5.6}$$

где $X, Y \in \mathbb{Z}_n^m$, $A = A(m, m)$ — некоторая обратимая матрица порядка m над кольцом $\mathbb{Z}_n$, $B = B(m, 1)$ — вектор-столбец над $\mathbb{Z}_n$.

Пусть $x_1 \dots x_l$ — некоторый открытый текст, где все $x_i \in \mathbb{Z}_n$. Разобьем данный текст на блоки равной длины m.

Каждый блок зашифруем с помощью аффинного преобразования (5.6). Соответственно расшифрование каждого блока длины m будет проходить по формуле:

$$X = A^{-1}(Y - B),$$

где $A^{-1} = A^{-1}(m, m)$ — обратная матрица для A. Заметим при этом, что квадратная матрица A над кольцом $\mathbb{Z}_n$ имеет обратную тогда и только тогда, когда $(det\ A, n) = 1$.

Таким образом, шифр Виженера является частным случаем аффинного блочного шифра при $A = E$, где E — единичная матрица. В случае же, когда A — некоторая обратимая матрица над $\mathbb{Z}_n$, а $B = \overline{0}$, то такой аффинный блочный шифр носит название *шифра Хилла*.

Аффинный блочный шифр значительно сложнее вскрыть по сравнению с шифром Виженера, однако свойство линейности рассматриваемого шифра значительно снижает его криптографическую стойкость.

Предположим, что криптоаналитику известны $m + 1$ блоков длины m открытого текста и соответствующие им блоки шифрованного текста: $X_0, X_1, \ldots, X_m, Y_0, Y_1, \ldots, Y_m$, полученных на одном ключе — A, B. Требуется найти этот ключ.

Рассмотрим систему матричных уравнений:

$$\begin{cases} AX_0 + B = Y_0 \\ AX_1 + B = Y_1 \\ \qquad \ldots \\ AX_m + B = Y_m \end{cases}$$

Вычитая уравнение $AX_0 + B = Y_0$ из уравнений $AX_i + B = Y_i$, $i = 1, 2, \ldots, m$, получим такую систему:

$$\begin{cases} A\widetilde{X}_1 = \widetilde{Y}_1 \\ \qquad \ldots \\ A\widetilde{X}_m = \widetilde{Y}_m, \end{cases}$$

где $\widetilde{X}_i = X_i - X_0$, $\widetilde{Y}_i = Y_i - Y_0$, $i = 1, 2, \ldots, m$. Из последней системы находится матрица A. После этого вектор-столбец B

определяется из любого матричного уравнения первой системы:

$$AX_i + B = Y_i.$$

5.2.6. Табличное гаммирование

Пусть все открытые и шифрованные тексты записываются в алфавите $A = \{a_1, \ldots, a_n\}$. Зафиксируем произвольным образом n биективных преобразований множества A : $\varphi_{a_1}, \ldots, \varphi_{a_n}$, где:

$$\varphi_{a_i} : A \to A, \quad i = 1, \ldots, n.$$

Пусть $x = x_1 \ldots x_l$ — некоторый открытый текст, где все $x_i \in A$, и $\gamma_1 \ldots \gamma_l$ — некоторое ключевое слово (гамма) в алфавите A, длина которого совпадает с длиной открытого текста x. Тогда шифрованный текст $y_1 \ldots y_l$ получается путем «наложения» ключевой последовательности $\gamma_1 \ldots \gamma_l$ на открытый текст $x_1 \ldots x_l$ следующим образом:

$$y_1 \ldots y_l = \varphi_{\gamma_1}(x_1) \ldots \varphi_{\gamma_l}(x_l).$$

Каждую подстановку φ_{a_i}, $i = 1, 2, \ldots, n$, можно представить в виде таблицы:

$$\varphi_{a_i} : \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ \varphi_{a_i}(a_1) & \varphi_{a_i}(a_2) & \ldots & \varphi_{a_i}(a_n) \end{pmatrix}.$$

Все подстановки $\varphi_{a_1}, \ldots, \varphi_{a_n}$ можно объединить в одну большую таблицу:

$$\left(\begin{array}{c|cccc} \varphi \diagdown A & a_1 & a_2 & \ldots & a_n \\ \hline \varphi_{a_1} : & \varphi_{a_1}(a_1) & \varphi_{a_1}(a_2) & \ldots & \varphi_{a_1}(a_n) \\ \varphi_{a_2} : & \varphi_{a_2}(a_1) & \varphi_{a_2}(a_2) & \ldots & \varphi_{a_2}(a_n) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \varphi_{a_n} : & \varphi_{a_n}(a_1) & \varphi_{a_n}(a_2) & \ldots & \varphi_{a_n}(a_n) \end{array}\right)$$

Если полученная таблица является латинским квадратом и определен способ получения ключевых последовательностей $\gamma_1 \ldots \gamma_l$ произвольной длины, то полученный шифр носит название *шифра табличного гаммирования.*

Обозначим через T^i циклическую подстановку на i позиций влево. Тогда если $\varphi_{a_i} = T^{i-1}$, $i = 1, 2, \ldots, n$, то полученная таблица:

$$\left(\begin{array}{c|ccccc} \varphi \diagdown A & a_1 & a_2 & \ldots & a_{n-1} & a_n \\ \hline \varphi_{a_1}: & a_1 & a_2 & \ldots & a_{n-1} & a_n \\ \varphi_{a_2}: & a_2 & a_3 & \ldots & a_n & a_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \varphi_{a_1}: & a_n & a_1 & \ldots & a_{n-2} & a_{n-1} \end{array}\right)$$

носит название *таблицы Виженера.*

5.2.7. Модульное гаммирование

Данный шифр является частным случаем табличного гаммирования. Пусть $A = \mathbb{Z}_n$. В данном случае в качестве биективных преобразований φ_0, φ_1,. . . ,φ_{n-1} множества $A = \mathbb{Z}_n$ будут выступать такие преобразования:

$$\varphi_i(x) = x + i \pmod{n}.$$

Рассмотрим некоторый открытый текст $x = x_1 \ldots x_l$, где все $x_i \in \mathbb{Z}_n$, и некоторую гамму $\gamma_1 \ldots \gamma_l$, где все $\gamma_i \in \mathbb{Z}_n$. Тогда шифрованный текст $y_1 \ldots y_l$ получается путем «наложения» гаммы $\gamma_1 \ldots \gamma_l$ на открытый текст $x_1 \ldots x_l$ следующим образом:

$$y_1 \ldots y_l = x_1 + \gamma_1 \ldots x_l + \gamma_l,$$

где «+» — бинарная операция из кольца вычетов $\mathbb{Z}_n$.

Расшифрование же шифртекста $y_1 \ldots y_l$ происходит по следующей формуле:

$$x_1 \ldots x_l = y_1 - \gamma_1 \ldots y_l - \gamma_l,$$

где «-» — бинарная операция из кольца вычетов $\mathbb{Z}_n$.

Шифр с данными правилами зашифрования и расшифрования носит название *шифра модульного гаммирования.*

Рассмотрим случай, когда $A = \mathbb{Z}_2$, т.е. все открытые тексты представляются в виде конечных двоичных последовательностей. Тогда при наложении гаммы $\gamma_1 \ldots \gamma_l$ на открытый текст

$x_1 \dots x_l$, где все x_i и $\gamma_i \in \mathbb{Z}_2$, используется операция сложения $\oplus$ по модулю 2:

$$y_1 \dots y_l = x_1 \oplus \gamma_1 \dots x_l \oplus \gamma_l.$$

Так как сложение и вычитание по модулю 2 совпадают, то расшифрование осуществляется по такой формуле:

$$x_1 \dots x_l = y_1 \oplus \gamma_1 \dots y_l \oplus \gamma_l.$$

Полученный шифр носит название *шифра Вернама.*

5.2.8. Шифр пропорциональной замены (шифр омофонов)

Как было отмечено, шифр простой замены легко раскрывается частотным криптоанализом. В целом, криптоанализ любого шифра невозможен без учета особенностей открытых текстов, которые подлежат шифрованию. Очень важной характеристикой открытых текстов является избыточность текста. Данная избыточность открытого текста проникает и в шифртекст, что является основной слабостью шифра.

Наиболее простыми характеристиками открытых текстов являются частотные характеристики букв, биграмм, триграмм, четырехграмм и т.д. Метод вскрытия шифра простой замены основан на том обстоятельстве, что с точностью до переобозначений частотные характеристики $m-$грамм открытого и шифрованного текстов одинаковы. При криптоанализе таких шифров используются априорные частотные характеристики открытого текста. Чем менее рельефно распределение символов открытого текста, тем труднее задача вскрытия шифра простой замены.

Чтобы скрыть частотные характеристики открытого текста, может быть использован шифр пропорциональной замены, при котором шифрование открытого текста происходит таким образом, чтобы каждый символ шифрованного текста имел бы (приблизительно) одинаковую частоту появления.

Пусть имеется источник сообщения с известными статистическими свойствами. Пусть $A = \{a_0, a_1, \dots, a_{n-1}\}$ — некоторый

алфавит, состоящий из n символов, с помощью которого записываются открытые тексты. Пусть также $p_0, p_1, \ldots, p_{n-1}$ — приближенные вероятности появления данных символов в открытом тексте. Если закон распределения:

$$\begin{array}{ll} A & a_0 \;\; a_1 \;\; \ldots \;\; a_{n-1} \\ P & p_0 \;\; p_1 \;\; \ldots \;\; p_{n-1}, \end{array}$$

где все $p_i > 0$, далек от «почти равномерного», то данную рельефность можно сгладить следующим образом.

Пусть B — некоторое конечное множество (состоящее из «достаточно» большого числа элементов), на котором зафиксируем некоторое разбиение $B = B_0 \cup B_1 \cup \ldots \cup B_{n-1}$ на непустые непересекающиеся подмножества $B_0, \ldots, B_{n-1}$ с тем условием, что:

$$\frac{|B_i|}{|B|} \approx p_i, \quad i = 0, 1, \ldots, n-1.$$

В данном случае множество B является алфавитом, с помощью которого будут записываться шифрованные тексты. Процесс шифрования будет осуществляться следующим образом: каждый символ открытого текста a_i заменяется на случайно (равновероятно) выбранный элемент из множества B_i. В этом случае при многократной замене символа a_i открытого текста элементами множества B_i каждый элемент алфавита B в шифрованном тексте будет использоваться примерно одинаковое число раз. Покажем это.

Пусть b — произвольный фиксированный элемент алфавита B. Тогда $b \in B_{i_0}$ для некоторого i_0. Оценим вероятность появления элемента b в шифрованном тексте. Заметим, что вероятность появления одного из символов множества B_i в шифрованном тексте совпадает с вероятностью появления символа a_i в открытом тексте, поэтому $P(B_i) = P(a_i) = p_i$ для любого $i = 0, 1, \ldots, n-1$. По формуле полной вероятности имеем:

$$P(b) = \sum_{i=0}^{n-1} P(B_i) \cdot P(b|B_i).$$

Поскольку $P(b|B_i) = 0$ для всех $i \neq i_0$, то:

$$P(b) = P(B_{i_0}) \cdot P(b|B_{i_0}) = p_{i_0} \cdot \frac{1}{|B_{i_0}|} \approx \frac{|B_{i_0}|}{|B|} \cdot \frac{1}{|B_{i_0}|} = \frac{1}{|B|}.$$

Таким образом, частоты появления элементов алфавита B в шифрованном тексте приблизительно одинаковы.

Процесс расшифрования не представляет трудностей: если b — очередной символ в шифрованном тексте, то сначала определяется множество B_i, к которому принадлежит символ b, после чего данный символ заменяется на символ a_i.

Приведем пример построения множества B, реализации разбиения на нем и процесс шифрования.

1. Для начала зафиксируем произвольное целое значение k, для которого выполнено следующее неравенство:

$$\min\{[k \cdot p_0],\ [k \cdot p_1], \ldots, [k \cdot p_{n-1}]\} \geqslant 1,$$

где [ ] - целая часть числа. Введем также следующие обозначения:

$$k_0 = [k \cdot p_0],\ k_1 = [k \cdot p_1], \ldots, k_{n-1} = [k \cdot p_{n-1}],$$
$$N = k_0 + k_1 + \ldots + k_{n-1}.$$

Заметим, что $N \leqslant k$, так как:

$$N = [k \cdot p_0] + [k \cdot p_1] + \ldots + [k \cdot p_{n-1}] \leqslant$$

$$\leqslant k \cdot p_0 + k \cdot p_1 + \ldots + k \cdot p_{n-1} = k \cdot (p_0 + p_1 + \cdots + p_{n-1}) = k.$$

Множество B будет состоять из чисел от 0 до $N-1$. Рассмотрим в данном множестве такие подмножества:

$$B_0 = \{0,\ 1, \ldots, k_0 - 1\},$$

$$B_1 = \{k_0,\ k_0 + 1, \ldots, k_0 + k_1 - 1\},$$

$$\ldots$$

$$B_{n-1} = \{k_0 + k_1 + \ldots + k_{n-2},\ k_0 + k_1 + \ldots k_{n-2} + 1, \ldots, N - 1\},$$

т.е. подмножества B_0, B_1,..., B_{n-1} образуют разбиение множества B. Заметим, что:

$$|B_0| = k_0,\ |B_1| = k_1, \ldots, |B_{n-1}| = k_{n-1}.$$

Далее процесс кодирования будет осуществляться следующим образом: каждый символ открытого текста a_i заменяется на случайно (равновероятно) выбранный элемент из множества B_i. После такого кодирования открытый текст X преобразуется в текст Y, символы которого принадлежат алфавиту $B = \{0,\ 1, \ldots, N-1\}$. Рассмотрим закон распределения появления символов алфавита B в тексте Y :

$$\begin{array}{cccccc} B & 0 & 1 & \ldots & N-1 \\ P & q_0 & q_1 & \ldots & q_{N-1}. \end{array}$$

Оценим значения вероятностей $q_0, q_1, \ldots, q_{N-1}$. Фиксируем некоторое целое число $i \in B$. Пусть B_{i_0} — такое множество, что $i \in B_{i_0}$. Тогда:

$$q_i = P(B_{i_0}) \cdot P(i|B_{i_0}) = p_{i_0} \cdot \frac{1}{k_{i_0}} = \frac{p_{i_0}}{[k \cdot p_{i_0}]}.$$

При «достаточно» большом значении k будет иметь место такое приближенное равенство: $q_i \approx 1/k \approx 1/N$. Таким образом, если взять «достаточно» большое k, то:

$$q_0 \approx \frac{1}{k},\ q_1 \approx \frac{1}{k}, \ldots, q_{N-1} \approx \frac{1}{k}.$$

Тем самым появление символов алфавита B в тексте Y становится «почти равномерным».

2. После того, как открытый текст X преобразован в Y, можно к Y применить тот или иной шифр, например, шифр простой замены, взломать который будет крайне сложно при достаточно большом k (например, при $k = 1000$ для русскоязычного или англоязычного текста).

Заметим, что пункты 1 и 2 в процессе шифрования можно объединить, параллельно кодируя символы алфавита A символами алфавита B и сразу же шифруя их, используя всего одно прохождение по открытому тексту X.

5.3. Шифры перестановки

Пусть $x = x_1 x_2 \ldots x_l$ — некоторый открытый текст, записанный в алфавите A, и σ — некоторая перестановка из симметрической группы S_l. Тогда шифрованный текст $y = y_1 y_2 \ldots y_l$

получается из x путем переупорядочивания букв в соответствии с перестановкой σ :

$$y_1 y_2 \dots y_l = x_{\sigma(1)} x_{\sigma(2)} \dots x_{\sigma(l)}.$$

Ключом шифрования является перестановка σ.

Расшифрование шифртекста y производится с помощью применения к нему обратной перестановки $\sigma^{-1} \in S_l$:

$$x_1 x_2 \dots x_l = y_{\sigma^{-1}(1)} y_{\sigma^{-1}(2)} \dots y_{\sigma^{-1}(l)}.$$

Такая модель шифра перестановки требует, чтобы длина открытого текста совпадала с длиной ключа. При больших значениях l это крайне неудобно с практической точки зрения. По этой причине чаще всего шифры перестановки используют ключ фиксированного размера s. При этом открытый текст разбивается на блоки длины s, к каждому из которых применяется ключевая перестановка.

5.3.1. Маршрутные перестановки

Данный вид перестановок использует прямоугольную таблицу, в которую открытый текст записывается по строкам, а шифрование происходит таким образом, что буквы из полученной таблицы извлекаются поочередно в соответствии с некоторым маршрутом.

Пример 5.1. Зашифруем фразу «применение маршрутной перестановки». Используем при этом таблицу 5×7 :

п	р	и	м	е	н	е
н	и	е		м	а	р
ш	р	у	т	н	о	й
	п	е	р	е	с	т
а	н	о	в	к	и	

В качестве маршрута (ключа) выберем направление, двигаясь с первого столбца к последнему таким образом:

↓ ↑ ... ↓

Получим такой шифртекст:

пнш анпририеуеоврт мемнекисоанерйт

Чтобы расшифровать данную криптограмму, необходимо записать ее в таблицу 5×7, двигаясь по ключевому маршруту.

Глава 6. Надежность шифров

6.1. Формальные модели шифров

Алгебраическая модель шифра. Рассмотрим алгебраическую модель шифра, предложенную К. Шенноном.

Пусть X, K, Y — конечные множества возможных открытых текстов, ключей и шифрованных текстов соответственно. Пусть также $E_k : X \to Y$ — правило зашифрования на ключе $k \in K$. Обозначим через E множество всех правил зашифрования $\{E_k \mid k \in K\}$, а через $E_k(X)$ — образ множества X при отображении $E_k : X \to Y$, т.е. $E_k(X) = \{E_k(x) \mid x \in X\}$. Пусть $D_k : E_k(X) \to X$ — правило расшифрования на ключе $k \in K$. Обозначим через D множество правил расшифрования $\{D_k \mid k \in K\}$.

Определение 6.1. Шифром (шифрсистемой) называется совокупность:

$$\Sigma_A = (X, K, Y, E, D),$$

для которой выполнены следующие свойства:

1) для любых $x \in X$, $k \in K$ выполнено $D_k(E_k(x)) = x$;

2) $Y = \bigcup_{k \in K} E_k(X)$.

Данное определение вводит математическую модель, отражающую основные свойства реальных шифров. Поэтому будем отождествлять реальный шифр с его моделью Σ_A, которая называется алгебраической моделью шифра.

Заметим, что условие 1 означает требование однозначности расшифрования, а условие 2 означает, что для любого элемента $y \in Y$ найдутся такие элементы $x \in X$ и $k \in K$, что y может быть представлен в следующем виде: $y = E_k(x)$.

Заметим также, что из условия 1 следует, что для любого $k \in K$ отображение E_k будет инъективным, т.е. для любых $x_1, x_2 \in X$, таких что $x_1 \neq x_2$, будет выполнено неравенство $E_k(x_1) \neq E_k(x_2)$ для любого $k \in K$ (если $E_k(x_1) = E_k(x_2)$ для некоторых $x_1 \neq x_2$ из множества X и некоторого $k \in K$, то из условия 1 следует, что $x_1 = D_k(E_k(x_1)) = D_k(E_k(x_2)) = x_2$, что противоречит условию $x_1 \neq x_2$).

Пусть $K = \{k_1, \ldots, k_m\}$, $X = \{x_1, \ldots, x_n\}$. Составим матрицу размера $m \times n$ над множеством Y, в которой строки пронумерованы элементами множества K, а столбцы — элементами множества X, а на пересечении строки с номером $k \in K$ и столбца с номером $x \in X$ поставим элемент $E_k(x)$:

$K \diagdown X$	x_1	$\ldots$	x_n
k_1	$E_{k_1}(x_1)$	$\ldots$	$E_{k_1}(x_n)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
k_m	$E_{k_m}(x_1)$	$\ldots$	$E_{k_m}(x_n)$

Полученная матрица A носит название *матрицы зашифрования шифра* Σ_A. Из условия 1 определения 6.1 следует, что в каждой строке матрицы A все элементы попарно различны, а из условия 2 — каждый элемент множества Y встречается в матрице A хотя бы один раз.

Произведение шифров. Пусть:

$$\Sigma_{A_1} = (X_1, K_1, Y_1, E^{(1)}, D^{(1)}), \quad \Sigma_{A_2} = (X_2, K_2, Y_2, E^{(2)}, D^{(2)}),$$

где $Y_1 = X_2$. *Произведением шифров* Σ_{A_1} и Σ_{A_2} называется шифр:

$$\Sigma_A = (X_1, K_1 \times K_2, Y_2, E, D),$$

для которого:

$$E_{(k_1,k_2)}(x) = E^{(2)}_{k_2}(E^{(1)}_{k_1}(x)),$$
$$D_{(k_1,k_2)}(y) = D^{(1)}_{k_1}(D^{(2)}_{k_2}(y)),$$

где $x \in X_1$, $y \in Y_2$, $(k_1, k_2) \in K_1 \times K_2$.

Вероятностная модель шифра. Пусть $P(X)$ и $P(K)$ — априорные распределения вероятностей соответственно на конечных множествах X и K, причем для любых $x \in X$ и $k \in K$

выполнены следующие неравенства: $P_X(x) > 0$, $P_K(k) > 0$, где $P_X(x) \in P(X)$, $P_K(k) \in P(K)$. Заметим, что:

$$\sum_{x \in X} P_X(x) = 1, \quad \sum_{k \in K} P_K(k) = 1.$$

Будем полагать, что выбор открытого текста и выбор ключа являются независимыми событиями (это равносильно тому, что распределения $P(X)$ и $P(K)$ независимы).

Определение 6.2. Под вероятностной моделью Σ_B будем понимать совокупность

$$\Sigma_B = (X, K, Y, E, D, P(X), P(K)),$$

где множества X, K и Y связаны условиями 1 и 2, которые указаны в определении 6.1.

Заметим, что распределения вероятностей $P(X)$ и $P(K)$ естественным образом индуцируют распределение вероятностей $P(Y) = \{P_Y(y) \mid y \in Y\}$ на множестве шифрованных текстов Y следующим образом:

$$P_Y(y) = \sum_{\substack{(x,k) \in X \times K \\ E_k(x) = y}} P_X(x) \cdot P_K(k). \tag{6.1}$$

Замечание 6.1. Заметим, что для произвольного $y \in Y$ выполнено неравенство $P_Y(y) > 0$. Это следует из определения 6.1, равенства (6.1) и того, что для любых $x \in X$ и $k \in K$ выполнено $P_X(x) > 0$ и $P_K(k) > 0$.

Обозначим через $K(x, y)$ множество таких ключей $k \in K$, для которых $E_k(x) = y$, т.е. $K(x, y) = \{k \in K \mid E_k(x) = y\}$. Условная вероятность $P_{Y|X}(y|x)$ определяется естественным образом:

$$P_{Y|X}(y|x) = \begin{cases} \sum_{k \in K(x,y)} P_K(k), & K(x, y) \neq \varnothing, \\ 0, & K(x, y) = \varnothing. \end{cases} \tag{6.2}$$

С помощью теоремы умножения вероятностей:

$$P(A \cdot B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

можно определить и условную вероятность $P_{X|Y}(x|y)$:

$$P_{X|Y}(x|y) = \frac{P_X(x) \cdot P_{Y|X}(y|x)}{P_Y(y)}.$$

В дальнейшем иногда вместо $P_X(x)$ будет записывать $P(x)$. Аналогично и с вероятностями $P_Y(y)$, $P_K(k)$, $P_{X|Y}(x|y)$, $P_{Y|X}(y|x)$.

6.2. Математические модели некоторых шифров

Пусть A и B — конечные алфавиты соответственно открытых и шифрованных текстов. Тогда в большинстве случаев множества X и Y представимы в следующем виде:

$$X = \bigcup_{i=1}^{L} A^i, \quad Y = \bigcup_{i=1}^{T} B^i.$$

Шифр простой замены. Пусть $X = \bigcup_{i=1}^{L} A^i$, $Y = \bigcup_{i=1}^{L} B^i$, причем $|A| = |B|$. Пусть также K представляет собой множество всех биективных отображений из A в B. Тогда для любого ключа $k \in K$, открытого текста $x = x_1 \dots x_l$ и шифрованного текста $y = y_1 \dots y_l$ правила зашифрования и расшифрования шифром простой замены определяются следующим образом:

$$y_1 \dots y_l = E_k(x_1 \dots x_l) = k(x_1) \dots k(x_l),$$

$$x_1 \dots x_l = D_k(y_1 \dots y_l) = k^{-1}(y_1) \dots k^{-1}(y_l),$$

где k^{-1} — биективное отображение из B в A, обратное к отображению k.

Шифр сдвига. Данный шифр является частным случаем шифра простой замены. Пусть $A = \{a_1, \dots, a_n\}$ — некоторый алфавит, состоящий из n символов. Сопоставим каждой букве данного алфавита ее порядковый номер, начиная с 0, т.е. алфавит A представим следующим образом: $A = \{0, 1, \dots, n-1\}$. Как и ранее, пусть $\mathbb{Z}_n$ — кольцо вычетов по модулю n. Пусть $X = Y = \bigcup_{i=1}^{L} Z_n^i$ и $K = \mathbb{Z}_n$. Тогда:

$$y_1 \dots y_l = E_k(x_1 \dots x_l) = x_1 + k \dots x_l + k,$$

$$x_1 \dots x_l = D_k(y_1 \dots y_l) = y_1 - k \dots y_l - k,$$

где «+» и «-» — операции в кольце $\mathbb{Z}_n$, т.е.:

$$y_i = x_i + k \pmod{n}, \quad x_i = y_i - k \pmod{n}.$$

Аффинный шифр. Пусть:

$$X = Y = \bigcup_{i=1}^{L} \mathbb{Z}_n^i, \quad K = \{(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n \mid (a, n) = 1\}.$$

Тогда правила зашифрования и расшифрования определяются следующим образом:

$$y_1 \dots y_l = E_k(x_1 \dots x_l) = ax_1 + b \dots ax_l + b,$$

$$x_1 \dots x_l = D_k(y_1 \dots y_l) = a^{-1}(y_1 + n - b) \dots a^{-1}(y_l + n - b).$$

Шифр замены с конечным ключом. Пусть A — конечный алфавит, в котором записываются открытые тексты, $B_1, \dots, B_s$ — конечные алфавиты, в которых записываются шифрованные тексты, причем:

$$|A| = |B_1| = \dots = |B_s|.$$

Множества X и Y определим следующим образом:

$$X = \bigcup_{i=1}^{L} (A^s)^i, \quad Y = \bigcup_{i=1}^{L} (B_1 \times \dots \times B_s)^i.$$

Обозначим через K_i множество всех биективных отображений из A в B_i, $i = 1, \dots, s$. Тогда множество ключей K будет иметь такой вид: $K = K_1 \times \dots \times K_s$.

Пусть $k = (k_1, \dots, k_s) \in K$. Тогда правила зашифрования и расшифрования на ключе k определяются следующим образом:

$$E_k(x_1 x_2 \dots x_s x_{s+1} x_{s+2} \dots x_{2s} \dots) =$$

$$= k_1(x_1) k_2(x_2) \dots k_s(x_s) k_1(x_{s+1}) k_2(x_{s+2}) \dots k_s(x_{2s}) \dots,$$

$$D_k(y_1 y_2 \dots y_s y_{s+1} y_{s+2} \dots y_{2s} \dots) =$$

$$= k_1^{-1}(y_1) k_2^{-1}(y_2) \dots k_s^{-1}(y_s) k_1^{-1}(y_{s+1}) k_2^{-1}(y_{s+2}) \dots k_s^{-1}(y_{2s}) \dots$$

Шифр перестановки. Пусть $X = Y = A^L$ и пусть $K = S_L$, где S_L — симметрическая группа всех подстановок множества $\{1, 2, \ldots, L\}$. Тогда для любого ключа $k \in K$, открытого текста $x = x_1 \ldots x_L$ и шифрованного текста $y = y_1 \ldots y_L$ правила зашифрования и расшифрования определяются следующим образом:

$$y_1 \ldots y_L = E_k(x_1 \ldots x_L) = x_{k(1)} \ldots x_{k(L)},$$

$$x_1 \ldots x_L = D_k(y_1 \ldots y_L) = y_{k^{-1}(1)} \ldots y_{k^{-1}(L)},$$

где k^{-1} — подстановка из S_L, обратная к k.

6.3. Ортогональные таблицы

Напомним несколько важных определений.

Латинским квадратом s-го порядка над множеством $Y = \{y_1, \ldots, y_s\}$ называется таблица размера $s \times s$, заполненная элементами множества Y таким образом, что в каждой строке и в каждом столбце каждый элемент встречается ровно один раз.

Две матрицы $A = (a_{ij})$ и $B = (b_{ij})$ над множеством $Y = \{y_1, \ldots, y_s\}$ называются *ортогональными*, если все упорядоченные пары (a_{ij}, b_{ij}) различны.

Ортогональной таблицей $OA(s, n)$ над s-элементным множеством $Y = \{y_1, \ldots, y_s\}$ называется матрица порядка $s^2 \times n$ над множеством Y с тем условием, что для любых двух столбцов данной матрицы каждая из пар $(y_i, y_j) \in Y \times Y$ встречается ровно один раз. Существование ортогональной таблицы $OA(s, n)$ над множеством Y эквивалентно существованию n попарно ортогональных квадратных матриц порядка s над множеством Y.

Хорошо известно, что если число s является степенью некоторого простого числа, то в этом случае существуют $s-1$ попарно ортогональных латинских квадрата, или, что тоже самое, $s+1$ ортогональных матриц: для этого достаточно рассмотреть многочлены $f_\alpha(x, y) = \alpha x + y$ над полем $GF(s)$ при ненулевых α.

Будем говорить, что матрица $A = A(s, n)$, $s \geqslant n$, над некоторым s-элементным множеством Y является *латинским пря-*

моугольником, если каждый столбец матрицы A является перестановкой элементов множества Y, причем в строках каждый элемент встречается не более одного раза.

6.4. Совершенные шифры

Определение 6.3. Шифр Σ_B называется совершенным (по Шеннону), если для любых $x \in X$ и $y \in Y$ выполняется следующее равенство:

$$P(x|y) = P(x).$$

Данное равенство означает, что апостериорные вероятности открытых текстов (вычисленные после получения шифрованного текста) совпадают с их априорными вероятностями. Другими словами, наличие шифрованного сообщения y не дает никакой дополнительной информации о том, на каком открытом тексте x было получено сообщение y.

В следующей лемме дадим эквивалентные условия совершенного шифра.

Лемма 6.1. Для произвольного шифра Σ_B следующие условия эквивалентны:

(i) для любых $x \in X$, $y \in Y$ выполнено равенство:

$$P(x|y) = P(x);$$

(ii) для любых $x \in X$, $y \in Y$ выполнено равенство:

$$P(y|x) = P(y);$$

(iii) для любых $x_1, x_2 \in X$ и $y \in Y$ выполнено равенство:

$$P(y|x_1) = P(y|x_2).$$

Доказательство. Равносильность условий (i) и (ii) следует из замечания 6.1, равенства $P(x)P(y|x) = P(y)P(x|y)$ и того, что для любых $x \in X$ и $y \in Y$ выполнены неравенства $P(x) > 0$ и $P(y) > 0$.

Очевидно, что из условия (ii) следует условие (iii).

Покажем, что из (iii) следует (ii). Зафиксируем произвольный элемент $y \in Y$. Условие (iii) означает, что вероятность

$P(y|x)$ не зависит от элементов $x \in X$. Обозначим данную вероятность через q. Тогда из формулы полной вероятности:

$$P(y) = \sum_{x \in X} P(x) \cdot P(y|x)$$

следует такое равенство:

$$P(y) = \sum_{x \in X} P(x) \cdot q = q \cdot \sum_{x \in X} P(x) = q.$$

Поскольку $q = P(y|x)$ для произвольного $x \in X$, то из справедливости условия (iii) следует (ii). $\square$

Покажем некоторые очевидные свойства совершенного шифра.

Лемма 6.2. Пусть Σ_B — совершенный шифр. Тогда для шифра Σ_B будут выполнены следующие свойства:

(i) для любых $x \in X$ и $y \in Y$ найдется такой ключ $k \in K$, что $E_k(x) = y$ (иными словами, для любых $x \in X$ и $y \in Y$ подмножество ключей $K(x, y)$ в K не является пустым);

(ii) для множеств X, Y и K справедливы следующие неравенства:

$$|X| \leqslant |Y| \leqslant |K|.$$

Доказательство. (i) Предположим противное. Пусть существуют такие $x_0 \in X$ и $y_0 \in Y$, что $K(x_0, y_0) = \varnothing$. Тогда из равенства (6.2) следует, что $P(y_0|x_0) = 0$. Так как шифр Σ_B является совершенным, то:

$$P(y_0|x_0) = P(y_0).$$

Следовательно, $P(y_0) = 0$. Противоречие.

(ii) Очевидно, что неравенство $|X| \leqslant |Y|$ выполнено для любого шифра.

Покажем неравенство $|Y| \leqslant |K|$. Зафиксируем произвольный элемент $x_0 \in X$ и определим функцию $\varphi : K \to Y$ следующим образом:

$$\varphi(k) = E_k(x_0), \quad k \in K. \tag{6.3}$$

Из пункта (i) следует, что данное отображение является сюръективным: для любого $y \in Y$ найдется такой $k \in K$, что $\varphi(k) = y$. Поэтому:

$$Im\ \varphi = \{E_k(x_0) \mid k \in K\} = Y.$$

Из данного равенства очевидным образом следует неравенство $|Y| \leqslant |K|$. $\square$

Замечание 6.2. Заметим, что условие (i) предыдущей леммы эквивалентно тому, что каждый элемент множества Y должен присутствовать во всех столбцах матрицы зашифрования совершенного шифра Σ_B.

Лемма 6.3. Пусть Σ_B — совершенный шифр. Тогда выполняется равенство $|K| = |Y|$ в том и только том случае, когда $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$.

Доказательство. Для начала заметим, что если A и B — некоторые конечные множества, причем $|A| = |B|$, то любое отображение $f : A \to B$ инъективно тогда и только тогда, когда оно сюръективно.

Пусть Σ_B — совершенный шифр и $|K| = |Y|$. Фиксируем произвольный элемент $x_0 \in X$. Из пункта (ii) леммы 6.2 следует, что отображение $\varphi : K \to Y$, определенное равенством (6.3), является сюръективным, а следовательно, и инъективным. Поэтому $|K(x_0, y)| = 1$ для любого $y \in Y$. В силу произвольности $x_0 \in X$ получаем, что $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$.

Обратно, пусть $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$. Тогда отображение $\varphi : K \to Y$ для любого фиксированного $x_0 \in X$ будет взаимно однозначным. Поэтому $|K| = |Y|$. $\square$

Теорема 6.1 (**достаточные условия совершенности шифра**). Пусть для шифра Σ_B выполнены следующие условия:

(i) $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$;

(ii) распределение вероятностей $P(K)$ является равномерным, т.е. $P(k) = \dfrac{1}{|K|}$ для любого $k \in K$.

Тогда шифр Σ_B является совершенным, причем распределение вероятностей $P(Y)$ будет являться равномерным и $|K| = |Y|$.

Доказательство. Пусть выполнены условия теоремы. Покажем, что в этом случае выполнен пункт (iii) леммы 6.1.

Из условия (i) следует, что для любых $x \in X$ и $y \in Y$ найдется, и притом единственный, элемент $k \in K$, такой, что $E_k(x) = y$. Поэтому, учитывая формулу (6.2) и пункт (ii), следует такое равенство:

$$P(y|x) = P(k) = \frac{1}{|K|}.$$

Таким образом, выполнен пункт (iii) леммы 6.1, из чего следует, что шифр Σ_B является совершенным.

Так как для любого $y \in Y$ выполнено равенство $P(y) = \dfrac{1}{|K|}$, то распределение вероятностей на множестве Y является равномерным. Равенство $|K| = |Y|$ следует из леммы 6.3. □

Очень часто на практике шифры обладают свойством $X = Y$. Данные шифры, следуя К. Шеннону, назовем *эндоморфными*. К. Шеннон полностью описал эндоморфные совершенные шифры с минимально возможным числом ключей. Из леммы 6.2 следует, что минимально возможное число ключей $|K|$ не меньше значения $|Y|$.

Теорема 6.2 (К. Шеннон). Пусть Σ_B — некоторый шифр, для которого выполнено равенство $|X| = |K| = |Y|$. Шифр Σ_B является совершенным тогда и только тогда, когда выполнены следующие условия:

(i) $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$;

(ii) распределение вероятностей $P(K)$ является равномерным, т.е. $P(k) = \dfrac{1}{|K|}$ для любого $k \in K$.

Доказательство. Пусть шифр Σ_B является совершенным и $|X| = |K| = |Y|$. Тогда из леммы 6.3 следует условие (i).

Покажем справедливость условия (ii). Запишем множество открытых текстов X следующим образом: $X = \{x_1, \ldots, x_N\}$. Зафиксируем произвольный элемент $y \in Y$. Из пункта (i) следует, что $|K(x_i, y)| = 1$ для всех $i = 1, 2, \ldots, N$. Пронумеруем множество ключей $K = \{k_1, \ldots, k_N\}$ таким образом, чтобы выполнялось равенство $E_{k_i}(x_i) = y$ для всех $i = 1, 2, \ldots, N$. Из равенства (6.2) будет следовать, что $P(y|x_i) = P(k_i)$ для любого $i = 1, 2, \ldots, N$. Так как шифр Σ_B является совершенным, то из последнего равенства и пункта (ii) леммы 6.1 будет следовать такое равенство:

$$P(k_i) = P(y|x_i) = P(y), \quad i = 1, 2, \ldots, N.$$

Тем самым условие (ii) доказано.

Доказательство в обратную сторону следует из теоремы 6.1. □

Пусть Σ_B — некоторый шифр, являющийся совершенным и для которого $|X| = |K| = |Y|$. Рассмотрим матрицу зашифрования данного шифра:

$K \diagdown X$	x_1	$\ldots$	x_N
k_1	$E_{k_1}(x_1)$	$\ldots$	$E_{k_1}(x_N)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
k_N	$E_{k_N}(x_1)$	$\ldots$	$E_{k_N}(x_N)$

Из условия 1 определения 6.1 произвольного шифра следует, что все строки данной матрицы являются некоторыми перестановками элементов множества Y. Из теоремы К. Шеннона следует, что все столбцы данной матрицы также являются некоторыми перестановками элементов множества Y. Поэтому матрица зашифрования данного шифра Σ_B будет являться латинским квадратом.

Пример 6.1. Пусть имеется несколько абонентов A, B, C и т.д. и некоторый набор открытых текстов x_1, x_2,..., x_N. Если множество открытых текстов $X = \{x_1, x_2, \ldots, x_N\}$ не очень большое (например, в качестве X могут выступать некоторые

секретные инструкции для военных частей), то организовать секретную переписку можно следующим образом. Зададим множество шифрованных сообщений $Y = \{1, 2, \ldots, N\}$ и множество ключей $K = \{k_1, k_2, \ldots, k_N\}$, причем пусть распределение вероятностей $P(K)$ является равномерным. Построим матрицу зашифрования таким образом, чтобы она являлась латинским квадратом, например:

$K \diagdown X$	x_1	x_2	x_3	$\ldots$	x_{N-1}	x_N
k_1	1	2	3	$\ldots$	$N-1$	N
k_2	2	3	4	$\ldots$	N	1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
k_N	N	1	2	$\ldots$	$N-2$	$N-1$

Предположим, что абоненты A, B, C и т.д. зафиксировали на данный момент времени ключ k_2. Если абонент A хочет передать абоненту B некоторое секретное сообщение, например x_3, то, шифруя его с помощью матрицы зашифрования, он получает шифрованное сообщение "4" и передает его абоненту B. С помощью матрицы зашифрования абонент B легко восстановит по ключу k_2 открытый текст x_3.

Заметим, что полученная шифрсистема будет являться совершенной.

Замечание 6.3. Заметим, что не только указанные в теореме шифры являются совершенными.

1. В качестве первого примера приведем неэндоморфный шифр с равномерным распределением ключей, который будет являться совершенным.

Пусть Σ_B — шифр, определенный множествами:

$$X = \{x_1, x_2\}, \quad K = \{k_1, k_2, k_3\}, \quad Y = \{y_1, y_2, y_3\},$$

и матрицей зашифрования

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_2	y_3
k_3	y_3	y_1

Данный шифр будет являться совершенным (теорема 6.1).

2. Приведем пример эндоморфного шифра с неравномерным распределением на множестве ключей.

Пусть Σ_B — шифр, определенный множествами:

$$X = \{x_1, x_2\}, \quad K = \{k_1, k_2, k_3\}, \quad Y = \{y_1, y_2\},$$

и матрицей зашифрования:

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_2	y_1
k_3	y_1	y_2

Пусть распределение вероятностей на множестве K имеет вид:

K	k_1	k_2	k_3
$P(K)$	$1/4$	$1/2$	$1/4$

Так как $P(y_1|x_1) = P(y_1|x_2)$ и $P(y_2|x_1) = P(y_2|x_2)$, то шифр Σ_B является совершенным (лемма 6.1).

3. В качестве еще одного примера приведем неэндоморфный шифр с неравномерным распределением ключей, для которого также не выполнено условие (i) теоремы Шеннона, но который будет являться совершенным.

Пусть Σ_B — шифр, определенный множествами:

$$X = \{x_1, x_2\}, \quad K = \{k_1, k_2, k_3, k_4\}, \quad Y = \{y_1, y_2, y_3\},$$

и матрицей зашифрования:

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_1	y_3
k_3	y_2	y_1
k_4	y_3	y_2

Найдем распределение $P(K)$, при котором шифр Σ_B будет являться совершенным. Воспользуемся эквивалентным условием совершенности шифра (iii) леммы 6.1, из которого следует,

что:

$$P(y_1|x_1) = P(y_1|x_2), \quad P(y_2|x_1) = P(y_2|x_2), \quad P(y_3|x_1) = P(y_3|x_2).$$

Обозначим $P(k_i) = p_i$, $i = 1, 2, 3, 4$. Применяя равенство (6.2) и $p_1 + p_2 + p_3 + p_4 = 1$, получим такую систему линейных относительно p_i уравнений:

$$\begin{cases} p_1 + p_2 + p_3 + p_4 = 1 \\ p_1 + p_2 = p_3 \\ p_3 = p_1 + p_4 \\ p_4 = p_2 \end{cases}$$

Расширенная матрица данной системы имеет такой вид:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{array}\right)$$

После приведения данной матрицы элементарными преобразованиями к ступенчатому виду получим такую матрицу:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Полагая теперь значение свободной переменной $p_4 = 1/4$, получим $p_1 = 1/8$, $p_2 = 1/4$, $p_3 = 3/8$. При таком, в частности, распределении $P(K)$ шифр Σ_B будет являться совершенным.

Наряду с теоремой Шеннона рассмотрим еще один критерий совершенных шифров в классе шифров с равномерным распределением вероятностей на множестве ключей K.

Теорема 6.3. Шифр Σ_B с равномерным распределением вероятностей $P(K)$ является совершенным тогда и только тогда, когда для любых $x_1, x_2 \in X$, $y \in Y$ выполнено равенство:

$$|K(x_1, y)| = |K(x_2, y)|.$$

Доказательство. Пусть шифр Σ_B является совершенным. Из леммы 6.1 следует, что для любых $x_1, x_2 \in X$ и $y \in Y$ выполнено равенство $P(y|x_1) = P(y|x_2)$. Так как распределение вероятностей на $P(K)$ является равномерным, то:

$$\frac{|K(x_1, y)|}{|K|} = P(y|x_1) = P(y|x_2) = \frac{|K(x_2, y)|}{|K|}.$$

Поэтому $|K(x_1, y)| = |K(x_2, y)|$ для любых $x_1, x_2 \in X$ и $y \in Y$.

Обратно, пусть для любых $x_1, x_2 \in X$, $y \in Y$ выполнено равенство $|K(x_1, y)| = |K(x_2, y)|$. Для начала покажем, что для любых $x \in X$ и $y \in Y$ найдется такой ключ $k \in K$, что $E_k(x) = y$. Предположим, что $K(x_0, y_0) = \varnothing$ для некоторых $x_0 \in X$, $y_0 \in Y$. Тогда $K(x, y_0) = \varnothing$ для любого $x \in X$. А это означает, что:

$$P(y_0) = \sum_{x \in X} P(x)P(y_0 \,|\, x) = 0,$$

что противоречит замечанию 6.1.

Осталось заметить, что:

$$P(y|x_1) = \frac{|K(x_1, y)|}{|K|} = \frac{|K(x_2, y)|}{|K|} = P(y|x_2).$$

Поэтому по лемме 6.1 следует, что шифр Σ_B является совершенным. □

Матрицы зашифрования шифров, рассматриваемых в теореме 6.3, обладают следующим свойством. Пусть $|Y| = m$. Каждый столбец матрицы зашифрования (порядка $|K| \times |X|$) содержит ровно n_1 экземпляров элемента y_1, n_2 — y_2,..., n_m — y_m, где $n_1 > 0$,...,$n_m > 0$ и $n_1 + \ldots + n_m = |K|$.

Пример 6.2. Пусть шифр Σ_B определен множествами:

$$X = \{x_1, x_2\}, \quad K = \{k_1, k_2, k_3, k_4\}, \quad Y = \{y_1, y_2, y_3\}$$

и матрицей зашифрования:

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_2	y_1
k_3	y_1	y_3
k_4	y_3	y_1

Тогда если распределение вероятностей на множестве K является равномерным, то шифр Σ_B является совершенным (теорема 6.3).

Следствие 6.1 (теоремы 6.3)**.** Пусть для шифра Σ_B выполнено равенство $|Y| = |K|$ и распределение вероятностей $P(K)$ является равномерным. Шифр Σ_B является совершенным тогда и только тогда, когда $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$.

Данное следствие утверждает, что если $|Y| = |K|$ и $P(K)$ равномерно, то совершенность шифра Σ_B эквивалентно тому, что матрица зашифрования шифра Σ_B является латинским прямоугольником.

Рассмотрим следующую задачу: по заданному множеству открытых текстов X_0 и множеству ключей K_0 с распределением вероятностей $P(K_0)$ (независимо от $P(X_0)$) однозначно определить, существует ли шифр:

$$\Sigma_B = (X_0, K_0, Y, E, D, P(X_0), P(K_0)),$$

являющийся совершенным. Таким образом, по заданным X_0, K_0, $P(K_0)$ требуется определить, найдутся ли такие Y, E, D, для которых шифр Σ_B являлся бы совершенным.

Теорема 6.4. Для заданных X, $|X| = n$, K, $|K| = m$, $P(K)$ существует совершенный шифр:

$$\Sigma_B = (X, K, Y, E, D, P(X), P(K))$$

тогда и только тогда, когда найдется такое натуральное число s и n разбиений множества K :

$$\begin{gathered} K = K_{11} \cup K_{12} \cup \ldots \cup K_{1s}, \quad K_{1i} \cap K_{1j} = \varnothing, \quad 1 \leqslant i < j \leqslant s, \\ K = K_{21} \cup K_{22} \cup \ldots \cup K_{2s}, \quad K_{2i} \cap K_{2j} = \varnothing, \quad 1 \leqslant i < j \leqslant s, \quad (6.4) \\ \ldots \\ K = K_{n1} \cup K_{n2} \cup \ldots \cup K_{ns}, \quad K_{ni} \cap K_{nj} = \varnothing, \quad 1 \leqslant i < j \leqslant s, \end{gathered}$$

для которых выполнены следующие условия:

1) $K_{it} \cap K_{jt} = \varnothing$, $1 \leqslant i < j \leqslant n$, $t = 1, \ldots, s$;

2) для любых $1 \leqslant i < j \leqslant n$, $t = 1, \ldots, s$ выполнено равенство:

$$\sum_{k \in K_{it}} P_K(k) = \sum_{k \in K_{jt}} P_K(k).$$

Доказательство. *Достаточность.* Пусть для X, K, $P(K)$, найдется такое s и n таких разбиений (6.4), для которых выполнены условия 1, 2. Пусть $Y = \{y_1, \ldots, y_s\}$ — некоторое множество шифрованных текстов, где s — число непустых частей в (6.4). Составим матрицу зашифрования размера $m \times n$, где строки пронумерованы элементами множества K, а столбцы — элементами множества X, следующим образом. В i-м столбце $(i = 1, \ldots, n)$ данной матрицы в строках, пронумерованных элементами множества K_{ij}, поставим элемент y_j, $j = 1, \ldots, s$. Условие 1 в этом случае гарантирует, что все правила зашифрования полученного шифра являются инъективными отображениями. А из условия 2 следует, что для любого $t = 1, \ldots, s$ и любых $1 \leqslant i < j \leqslant n$ будут выполнены равенства:

$$P_{Y|X}(y_t|x_i) = \sum_{k \in K_{it}} P_K(k) = \sum_{k \in K_{jt}} P_K(k) = P_{Y|X}(y_t|x_j).$$

Поэтому, учитывая лемму 6.1, полученный шифр будет являться совершенным.

Необходимость. Пусть для заданных X, K, $P(K)$ существует совершенный шифр Σ_B со множеством шифрованных текстов $Y = \{y_1, \ldots, y_s\}$. Обозначим для данного шифра:

$$K_{it} = \{k \in K \mid E_k(x_i) = y_t\}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, s.$$

Понятно, что:

$$P_{Y|X}(y_t|x_i) = \sum_{k \in K_{it}} P_K(k).$$

Из лемм 6.1 и 6.2 следует, что для множеств K_{it} будут выполнены равенства (6.4) и условия 1, 2. □

Следствие 6.2. Пусть для заданных X, K, $P(K)$ существует совершенный шифр. Тогда для любого множества открытых текстов $\widetilde{X}$, $|\widetilde{X}| \leqslant |X|$, и для заданных K, $P(K)$ существует совершенный шифр.

Следствие 6.3. Для заданных X, $|X| = n$, K, $P(K)$, Y, $|Y| = s$, существует совершенный шифр:

$$\Sigma_B = (X, K, Y, E, D, P(X), P(K))$$

тогда и только тогда, когда найдется n таких разбиений (6.4), для которых выполнены условия 1, 2 предыдущей теоремы

Следствие 6.4. Для заданных Y, $|Y| = s$, K, $P(K)$, существует совершенный шифр Σ_B тогда и только тогда, когда найдется такое n и такие разбиения (6.4), для которых выполнены условия 1, 2 предыдущей теоремы.

Пример 6.3. Пусть $X = \{x_1, x_2\}$, $K = \{k_1, k_2, k_3, k_4\}$ и распределение вероятностей на множестве K имеет вид:

K	k_1	k_2	k_3	k_4
$P(K)$	1/8	1/4	3/8	1/4

В этом случае можно построить два разбиения множества K вида:

$$K = \{k_1, k_2\} \cup \{k_3\} \cup \{k_4\},$$

$$K = \{k_3\} \cup \{k_1, k_4\} \cup \{k_2\},$$

где $\{k_1, k_2\} \cap \{k_3\} = \{k_3\} \cap \{k_1, k_4\} = \{k_4\} \cap \{k_2\} = \varnothing$. При этом будут выполнены равенства:

$$P_K(k_1) + P_K(k_2) = P_K(k_3),$$

$$P_K(k_3) = P_K(k_1) + P_K(k_4),$$

$$P_K(k_4) = P_K(k_2).$$

По теореме 6.4 для данных X, K, $P(K)$ можно построить совершенный шифр. Пусть $Y = \{y_1, y_2, y_3\}$. Составим матрицу зашифрования следующим образом:

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_1	y_3
k_3	y_2	y_1
k_4	y_3	y_2

Тогда полученный шифр будет являться совершенным.

Лемма 6.4. Для заданных X и Y можно построить совершенный шифр Σ_B тогда и только тогда, когда $|X| \leqslant |Y|$.

Доказательство. Если шифр Σ_B является совершенным, то неравенство $|X| \leqslant |Y|$ следует из леммы 6.2.

Обратно, пусть для X и Y выполнено неравенство $|X| \leqslant |Y|$. Обозначим $m = |Y|$, $n = |X|$. Составим матрицу A порядка $m \times n$ над множеством Y следующим образом: в каждом столбце матрицы A каждый элемент множества Y встречается ровно один раз, а в каждой строке нет повторяющихся элементов. Пусть матрица A будет матрицей зашифрования для шифра Σ_B, а распределение вероятностей на множестве K равномерно, $|K| = m$. Тогда из теоремы 6.1 следует, что шифр Σ_B является совершенным. □

6.5. $(k|y)$-совершенные шифры

Нам понадобится следующее несложное утверждение.

Лемма 6.5. Для произвольного шифра Σ_B следующие условия эквивалентны:

(i) $|X| = |Y|$;

(ii) для любой пары $(k, y) \in K \times Y$ найдется, и притом единственный, $x \in X$, такой что $E_k(x) = y$;

(iii) для любой пары $(k, y) \in K \times Y$ найдется $x \in X$, такой что $E_k(x) = y$.

Доказательство. $(i) \Rightarrow (ii)$. Пусть выполнено условие (i). Зафиксируем произвольное значение $k_0 \in K$. Так как отображение $E_{k_0} : X \to Y$ является инъекцией и $|X| = |Y|$, то E_{k_0} является биективным отображением. Поэтому для любого значения $y \in Y$ найдется, и притом единственный, элемент $x \in X$, что $E_{k_0}(x) = y$. В силу произвольности $k_0 \in K$ получаем справедливость условия (ii).

$(ii) \Rightarrow (iii)$ очевидно.

$(iii) \Rightarrow (i)$ Пусть выполнено условие (iii). Зафиксируем произвольным образом $k_0 \in K$. Из условия (iii) следует, что отображение $E_{k_0} : X \to Y$ является сюръекцией. А так как в силу определения понятия шифра отображение E_{k_0} является инъекцией, то E_{k_0} будет являться биективным отображением. Поэтому $|X| = |Y|$. $\square$

Определение 6.4. Шифр Σ_B называется $(k|y)$-совершенным (совершенным по ключу), если для любых $k \in K$ и $y \in Y$ выполняется следующее равенство:

$$P(k|y) = P(k).$$

Данное равенство означает, что наличие шифрованного сообщения y не дает никакой дополнительной информации о том, на каком ключе было получено сообщение y.

Обозначим $X(k, y) = \{x \in X \mid E_k(x) = y\}$. Заметим, что множество $X(k, y)$ либо пусто, либо одноэлементно.

Далее нам понадобятся следующие формулы:

$$P(y|k) = \begin{cases} P(x), & X(k, y) = \{x\}, \\ 0, & X(k, y) = \varnothing, \end{cases} \tag{6.5}$$

$$P(k)P(y|k) = P(y)P(k|y).$$

Доказательство следующей леммы аналогично доказательству леммы 6.1.

Лемма 6.6. Для произвольного шифра Σ_B следующие условия эквивалентны:

(i) для любых $k \in K$, $y \in Y$ выполнено равенство:

$$P(k|y) = P(k);$$

(ii) для любых $k \in K$, $y \in Y$ выполнено равенство:

$$P(y|k) = P(y);$$

(iii) для любых $k_1, k_2 \in K$ и $y \in Y$ выполнено равенство:

$$P(y|k_1) = P(y|k_2).$$

Лемма 6.7. Пусть шифр Σ_B является $(k|y)$-совершенным. Тогда $|X| = |Y|$.

Доказательство. Покажем, что если некоторый шифр Σ_B является $(k|y)$-совершенным, то для данного шифра будет выполнено условие (iii) леммы 6.5. Предположим, что это не так. Пусть для некоторой пары $(k_0, y_0) \in K \times Y$ выполнено равенство $X(k_0, y_0) = \varnothing$. Тогда из леммы 6.6 и равенства (6.5) будет следовать такое равенство:

$$P(y_0) = P(y_0|k_0) = 0.$$

Данное равенство противоречит тому, что для любого $y \in Y$ выполнено неравенство $P(y) > 0$.

Таким образом, для шифра Σ_B выполнено условие (iii) леммы 6.5, а следовательно, и условие (i). □

Теорема 6.5 (достаточные условия $(k|y)$-совершенного шифра). Пусть для шифра Σ_B выполнены следующие условия:

(i) $|X| = |Y|$;

(ii) распределение вероятностей $P(X)$ является равномерным.

Тогда шифр Σ_B является $(k|y)$-совершенным, причем распределение вероятностей $P(Y)$ будет являться равномерным.

Доказательство. Пусть выполнены условия теоремы. Тогда из условия (i) данной теоремы и леммы 6.5 следует, что для любой пары $(k, y) \in K \times Y$ найдется, и притом единственный, $x \in X$, такой что $E_k(x) = y$. Поэтому, учитывая условие (ii) и равенство (6.5), будет выполнено такое равенство:

$$P(y|k) = P(x) = \frac{1}{|X|} = \frac{1}{|Y|}.$$

Следовательно, для любых $y_1, y_2 \in Y$, $k_1, k_2 \in K$ справедливо следующее равенство:

$$P(y_1|k_1) = \frac{1}{|Y|} = P(y_2|k_2).$$

Таким образом, выполнено условие (iii) леммы 6.6, что доказывает $(k|y)$-совершенность шифра Σ_B.

Из той же леммы следует, что для любых $y \in Y$, $k \in K$ будет выполнено равенство $P(y|k) = P(y)$. Поэтому:

$$P(y) = P(y|k) = \frac{1}{|Y|}. \qquad \square$$

Пример 6.4. Пусть для шифра Σ_B :

$$X = \{x_1, x_2, x_3\}, \quad Y = \{y_1, y_2, y_3\}, \quad K = \{k_1, k_2\},$$

матрица зашифрования имеет вид:

$K \diagdown X$	x_1	x_2	x_3
k_1	y_1	y_2	y_3
k_2	y_2	y_3	y_1

,

распределение вероятностей $P(X)$ равномерно, $P(K)$ произвольно. Тогда шифр Σ_B является $(k|y)$-совершенным (теорема 6.5), но не является совершенным (лемма 6.2).

Пример 6.5. Приведем пример $(k|y)$-совершенного шифра с неравномерным распределением на множестве X.

Пусть для шифра Σ_B :

$$X = \{x_1, x_2, x_3\}, \quad Y = \{y_1, y_2, y_3\}, \quad K = \{k_1, k_2\},$$

матрица зашифрования имеет вид:

$K \diagdown X$	x_1	x_2	x_3
k_1	y_1	y_2	y_3
k_2	y_3	y_2	y_1

,

$P(K)$ произвольно, распределение вероятностей $P(X)$ имеет следующий вид:

X	x_1	x_2	x_3
$P(X)$	$1/4$	$1/2$	$1/4$.

Тогда шифр Σ_B является $(k|y)$-совершенным (лемма 6.6), но не является совершенным (лемма 6.2).

Лемма 6.8 (необходимые условия одновременно совершенных и $(k|y)$-совершенных шифров). Пусть шифр

Σ_B является одновременно совершенным и $(k|y)$-совершенным. Тогда для шифра Σ_B выполнены следующие условия:

1) для любых $x \in X$ и $y \in Y$ найдется такой ключ $k \in K$, что $E_k(x) = y$;

2) $|X| = |Y| \leqslant |K|$;

3) распределение вероятностей $P(X)$ равномерно;

4) распределение вероятностей $P(Y)$ равномерно.

Доказательство. Условия 1 и 2 следуют из лемм 6.2 и 6.7.

Покажем справедливость условий 3 и 4. Зафиксируем произвольным образом $y \in Y$. Из условия 1 следует, что шифртекст y присутствует во всех столбцах матрицы зашифрования шифра Σ_B. Пронумеруем элементы множеств $K = \{k_1, \ldots, k_m\}$, $X = \{x_1, \ldots, x_n\}$, $n \leqslant m$, таким образом, чтобы выполнялось равенство $E_{k_i}(x_i) = y$, $i = 1, \ldots, n$. Тогда:

$$P(x_i) = P(y|k_i) = P(y), \ i = 1, \ldots, n.$$

Поэтому $P(X)$ имеет равномерное распределение. При этом в силу произвольности выбора $y \in Y$, имеем:

$$P(y) = P(x) = \frac{1}{|X|} = \frac{1}{|Y|}. \qquad \square$$

Теорема 6.6. Пусть для шифра Σ_B выполнено равенство $|X| = |Y|$ и распределения вероятностей $P(X)$ и $P(K)$ равномерны. Шифр Σ_B является совершенным и $(k|y)$-совершенным тогда и только тогда, когда для любых $x_1, x_2 \in X$, $y \in Y$ выполнено равенство $|K(x_1, y)| = |K(x_2, y)|$.

Доказательство следует из теорем 6.3 и 6.5. $\square$

Теорема 6.7. Пусть для шифра Σ_B выполнены равенства $|X| = |Y| = |K|$. Шифр Σ_B является совершенным и $(k|y)$-совершенным тогда и только тогда, когда выполнены следующие условия:

(i) $|K(x, y)| = 1$ для любых $x \in X$ и $y \in Y$;

(ii) распределение вероятностей $P(K)$ является равномерным;

(iii) распределение вероятностей $P(X)$ является равномерным.

Доказательство следует из теорем 6.2, 6.5 и леммы 6.8. □

Лемма 6.9. Для заданных X, Y существует $(k|y)$-совершенный шифр:

$$\Sigma_B = (X, K, Y, E, D, P_X, P_K)$$

тогда и только тогда, когда выполнено равенство $|X| = |Y|$.

Доказательство. *Необходимое* условие $(k|y)$-совершенного шифра следует из леммы 6.7.

Достаточность. Пусть $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$. Понятно, что в этом случае для любого натурального $m > 1$ найдутся такие перестановки $\sigma_1, \ldots, \sigma_m \in S_n$, для которых выполнены следующие равенства:

$$P_X(x_{\sigma_i(s)}) = P_X(x_{\sigma_j(s)}), \quad 1 \leqslant i < j \leqslant m, \quad s = 1, \ldots, n. \quad (6.6)$$

Например, в качестве таких перестановок можно взять тождественные перестановки. Пусть для некоторого фиксированного m перестановки $\sigma_1, \ldots, \sigma_m \in S_n$ обладают условием (6.6). Пусть $K = \{k_1, \ldots, k_m\}$ — некоторое множество ключей. Составим матрицу зашифрования размера $m \times n$, где строки пронумерованы элементами множества K, а столбцы — элементами множества X, следующим образом: на позицию $(i, \sigma_i(s))$ поставим шифртекст y_s, $i = 1, \ldots, m$, $s = 1, \ldots, n$. Пусть $1 \leqslant i < j \leqslant m$, $1 \leqslant s \leqslant n$. Так как $X(k_i, y_s) = \{x_{\sigma_i(s)}\}$, то:

$$P_{Y|K}(y_s|k_i) = P_X(x_{\sigma_i(s)}) = P_X(x_{\sigma_j(s)}) = P_{Y|K}(y_s|k_j).$$

Поэтому из леммы 6.6 следует, что шифр Σ_B является $(k|y)$-совершенным. □

Рассмотрим следующую задачу (с учетом леммы 6.8): по заданному множеству шифрованных текстов Y_0, множеству открытых текстов X_0 с равномерным распределением вероятностей $P(X_0)$, множеству ключей K_0 с распределением вероятностей $P(K_0)$ однозначно определить, существует ли шифр:

$$\Sigma_B = (X_0, K_0, Y_0, E, D, P(X_0), P(K_0)),$$

являющийся одновременно совершенным и $(k|y)$-совершенным.

Теорема 6.8. Для заданных $Y = \{y_1, \ldots, y_n\}$, $X = \{x_1, \ldots, x_n\}$ с равномерным распределением $P(X)$, K с распределением вероятностей $P(K)$ существует одновременно совершенный и $(k|y)$-совершенный шифр

$$\Sigma_B = (X, K, Y, E, D, P(X), P(K))$$

тогда и только тогда, когда найдется такая матрица $A = A(K)$ порядка $n \times n$, каждый элемент которой является непустым подмножеством в K, для которой выполнены следующие условия:

1) каждая строка и каждый столбец матрицы A является разбиением множества K на непересекающиеся подмножества;

2) для любых $i = 1, \ldots, n$, $j = 1, \ldots, n$ выполнено равенство:

$$\sum_{k \in A_{ij}} P(k) = \frac{1}{n}.$$

Доказательство следует из теорем 6.4 и 6.5.

Пример 6.6. Пусть $X = \{x_1, x_2, x_3\}$, распределение вероятностей $P(X)$ равномерно, $Y = \{y_1, y_2, y_3\}$, $K = \{k_1, k_2, k_3, k_4, k_5\}$ и распределение вероятностей на множестве K имеет вид:

K	k_1	k_2	k_3	k_4	k_5
$P(K)$	$1/15$	$4/15$	$1/9$	$2/9$	$1/3$

В этом случае матрицу A из теоремы 6.8 можно построить следующим образом:

$\{k_1, k_2\}$	$\{k_3, k_4\}$	$\{k_5\}$
$\{k_3, k_4\}$	$\{k_5\}$	$\{k_1, k_2\}$
$\{k_5\}$	$\{k_1, k_2\}$	$\{k_3, k_4\}$

Составим матрицу зашифрования следующим образом:

$K \diagdown X$	x_1	x_2	x_3
k_1	y_1	y_3	y_2
k_2	y_1	y_3	y_2
k_3	y_2	y_1	y_3
k_4	y_2	y_1	y_3
k_5	y_3	y_2	y_1

Тогда полученный шифр будет являться совершенным и $(k|y)$-совершенным.

6.6. Математические модели шифра замены с ограниченным и неограниченным ключом

Определенная ранее вероятностная модель шифра Σ_B позволяет рассматривать в качестве множества открытых текстов X лишь слова (тексты) в некотором конечном алфавите A, длины которых ограничены некоторой заранее определенной константой. Такая модель не предполагает рассмотрения (бесконечного) множества всех конечных слов в алфавите A. То же самое касается и числа ключей. Шифр гаммирования, например, имеет бесконечное множество ключей, которыми служат всевозможные отрезки гаммы конечной длины.

Рассмотрим математическую модель шифра замены, предложенную А.Ю. Зубовым [22].

Пусть A и B — некоторые конечные множества. Будем считать, что открытые и шифрованные тексты являются словами в алфавитах A и B соответственно, т.е. $X \subset A^*$, $Y \subset B^*$, где A^* и B^* — множество всех слов конечной длины соответственно в алфавитах A и B.

Перед зашифрованием открытый текст $x \in X$ предварительно представляется в виде последовательности подслов, называемых *шифрвеличинами*:

$$x = \underbrace{\ldots}_{u_1}\underbrace{\ldots}_{u_2}\cdots\underbrace{\ldots}_{u_l}$$

В процессе зашифрования шифрвеличины заменяются некоторыми из эквивалентов в шифртексте, которые будем называть *шифробозначениями*.

Пусть U — конечное множество возможных шифрвеличин, а V — конечное множество возможных шифробозначений. Данные множества должны удовлетворять следующему свойству: любые тексты $x \in X$, $y \in Y$ представимы словами из U^* и V^* соответственно. Требование однозначности расшифрования влечет такое неравенство: $|U| \leqslant |V|$.

Пусть также имеются r $(r > 1)$ инъективных отображений из U в V. Пронумеруем данные отображения: E_1, E_2,..., E_r, которые можно задать в виде таблиц:

$$E_1 : \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ v_1^1 & v_2^1 & \dots & v_n^1 \end{pmatrix}, \dots, \; E_r : \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ v_1^r & v_2^r & \dots & v_n^r \end{pmatrix}.$$

Обозначим:

$$\mathbb{N}_r = \{1, 2, \dots, r\}.$$

Процесс зашифрования открытого текста $x = u_1 \dots u_l$ заключается в замене каждой шифрвеличины u_i на шифробозначение v_i, $i = 1, \dots, l$, в соответствии с одним из r инъективных отображений E_j, $j \in \mathbb{N}_r$. Отображения E_j, $j \in \mathbb{N}_r$, будем называть *простыми заменами*. Заметим, что максимальное число данных простых замен не превышает числа размещений из $|V|$ по $|U|$.

Обозначим через D_j отображение из $E_j(U)$ в U, $j \in \mathbb{N}_r$.

Определение 6.5. *Опорным шифром* шифра замены назовем совокупность:

$$\Sigma = (U, \mathbb{N}_r, V, E, D),$$

для которой выполнены следующие свойства:

1) для любых $u \in U$, $j \in \mathbb{N}_r$ выполнено $D_j(E_j(u)) = u$;

2) $V = \bigcup_{j \in \mathbb{N}_r} E_j(U)$.

Опорный шифр определяет способ зашифрования отдельных шифрвеличин. Для шифрования последовательностей шифрвеличин требуется ввести понятие степени опорного шифра.

Определение 6.6. *l-й степенью опорного шифра* Σ назовем совокупность:

$$\Sigma^l = (U^l, \mathbb{N}_r^l, V^l, E^{(l)}, D^{(l)}),$$

где $U^l, \mathbb{N}_r^l, V^l$ — декартовы степени соответствующих множеств U, $\mathbb{N}_r$, V; множество $E^{(l)}$ состоит из отображений:

$$E_{\bar{\jmath}} : U^l \to V^l, \quad \bar{\jmath} \in \mathbb{N}_r^l,$$

таких, что для любых $\overline{u} = u_1 \dots u_l \in U^l$, $\overline{j} = j_1 \dots j_l \in \mathbb{N}_r^l$ выполнено равенство:

$$E_{\overline{j}}(\overline{u}) = E_{j_1}(u_1) \dots E_{j_l}(u_l) = v_1 \dots v_l \in V^l;$$

множество $D^{(l)}$ состоит из отображений:

$$D_{\overline{j}} : E_{\overline{j}}(U^l) \to U^l, \quad \overline{j} \in \mathbb{N}_r^l,$$

таких, что для любых $\overline{v} = v_1 \dots v_l \in V^l$, $\overline{j} = j_1 \dots j_l \in \mathbb{N}_r^l$ выполнено равенство:

$$D_{\overline{j}}(\overline{v}) = D_{j_1}(v_1) \dots D_{j_l}(v_l) = u_1 \dots u_l \in U^l.$$

Заметим, что при $l = 1$ шифры Σ^l и Σ совпадают.

Теперь поясним, как строится ключевой поток, т.е. последовательность $j_1 \dots j_l \in \mathbb{N}_r^l$.

Случайный и детерминированный генераторы ключевого потока. Имеется два способа построения такой последовательности. В первом случае она строится случайно. Назовем *случайным генератором ключевого потока* устройство, которое вырабатывает случайный ключевой поток. Формально его можно представить отображением $\psi_c : \mathbb{N} \to \mathbb{N}_r^*$, ставящим в соответствие натуральному числу $l \in \mathbb{N}$ последовательность из l испытаний некоторой случайной величины, принимающей значения из $\mathbb{N}_r$ с ненулевыми вероятностями.

Во втором случае шифр имеет априорно заданное конечное множество ключей K. Каждому ключу $k \in K$ и натуральному числу $l \in \mathbb{N}$ однозначно ставится в соответствие ключевой поток $j_1 \dots j_l \in \mathbb{N}_r^l$. Данный поток вырабатывается детерминированным генератором ключевого потока.

Определение 6.7. *Детерминированным генератором ключевого потока* назовем отображение:

$$\psi_d : K \times \mathbb{N} \to \mathbb{N}_r^*,$$

такое, что для любых $k \in K$, $l \in \mathbb{N}$ следует, что:

$$\psi_d(k, l) = j_1 \dots j_l \in \mathbb{N}_r^l,$$

причем $\{\psi_d(k, 1) \mid k \in K\} = \mathbb{N}_r$. Ключевой поток $j_1 \dots j_l$ назовем потоком, отвечающем ключу k и числу l.

Случайный выбор ключей индуцирует распределение вероятностей на множестве ключевых потоков $\mathbb{N}_r^l$, $l \in \mathbb{N}$.

Пусть для любого $l \in \mathbb{N}$

$$P(\mathbb{N}_r^l) = \{P_{\mathbb{N}_r^l}(\overline{\jmath}),\ \overline{\jmath} \in \mathbb{N}_r^l\},$$

$$P(U^l) = \{P_{U^l}(\overline{u}),\ \overline{u} \in U^l\}$$

— априорные распределения вероятностей соответственно на множествах $\mathbb{N}_r^l$ и U^l, которые полагаются независимыми.

Замечание 6.4. Отметим такой важный момент. В ряде случаев не всякое слово длины l в алфавите U может появиться в открытом тексте. Например, в тексте на русском языке не может встретиться биграмма ьь. Поэтому обозначим через $U^{(l)}$ подмножество всех таких слов во множестве U^l, появление которых в открытом тексте имеет ненулевую вероятность:

$$U^{(l)} = \{\overline{u} \in U^l \mid P_{U^l}(\overline{u}) > 0\}.$$

То же самое можно сказать и о ключевых потоках, поскольку детерминированный генератор вырабатывает не всякую ключевую последовательность из $\mathbb{N}_r^l$, так как:

$$|\mathbb{N}_r^l| = r^l, \quad |\{\psi_d(k,l) \mid k \in K\}| \leqslant |K|.$$

Обозначим через $\mathbb{N}_r^{(l)}$ множество возможных ключевых потоков длины l, вырабатываемых детерминированным генератором:

$$\mathbb{N}_r^{(l)} = \{\psi_d(k,l) \mid k \in K\} = \{\overline{\jmath} \in \mathbb{N}_r^l \mid \exists k \in K :\ \psi_d(k,l) = \overline{\jmath}\}.$$

Заметим, что в случае случайного генератора будет иметь место равенство: $\mathbb{N}_r^{(l)} = \mathbb{N}_r^l$.

С учетом всего сказанного, полагаем:

$$V^{(l)} = \bigcup_{\overline{\jmath} \in \mathbb{N}_r^{(l)}} E_{\overline{\jmath}}(U^{(l)}).$$

Шифр замены с неограниченным ключом. Обозначим через $\Sigma_H^{(l)}$ следующую совокупность величин:

$$\Sigma_H^{(l)} = (U^{(l)}, \mathbb{N}_r^l, V^{(l)}, E^{(l)}, D^{(l)}, P(U^{(l)}), P(\mathbb{N}_r^l)), \qquad (6.7)$$

где распределение вероятностей $P(\mathbb{N}_r)$ индуцирует распределение вероятностей $P(\mathbb{N}_r^l)$ следующим образом:

$$P(\bar{\jmath}) = P(j_1 \dots j_l) = \prod_{i=1}^{l} P(j_i), \quad \bar{\jmath} \in \mathbb{N}_r^l.$$

Определение 6.8. *Шифром замены с неограниченным ключом* назовем семейство:

$$\Sigma_H = (\Sigma_H^{(l)},\ l \in \mathbb{N};\ \psi_c),$$

где ψ_c — случайный генератор ключевого потока. При этом совокупность (6.7) будем называть l-ым опорным шифром шифра Σ_H.

Шифр замены с ограниченным ключом. Построим теперь модель шифра замены с ограниченным ключом. Пусть имеется некоторое конечное множество ключей K, причем априорное распределение $P(K)$ не содержит нулевых вероятностей. Распределение $P(K)$ индуцирует распределение $P(\mathbb{N}_r^{(l)})$, $l \in \mathbb{N}$, следующим образом.

Пусть $\bar{\jmath} = j_1 \dots j_l \in \mathbb{N}_r^{(l)}$. Обозначим:

$$K_l(\bar{\jmath}) = \{k \in K \mid \psi_d(k, l) = \bar{\jmath}\},$$

где ψ_d — детерминированный генератор ключевого потока. Тогда значения вероятностей из $P(\mathbb{N}_r^{(l)})$ вычисляются по следующей формуле:

$$P_{\mathbb{N}_r^{(l)}}(\bar{\jmath}) = \sum_{k \in K_l(\bar{\jmath})} P_K(k). \tag{6.8}$$

Обозначим через $\Sigma_O^{(l)}$ следующую совокупность величин:

$$\Sigma_O^{(l)} = (U^{(l)}, \mathbb{N}_r^{(l)}, V^{(l)}, E^{(l)}, D^{(l)}, P(U^{(l)}), P(\mathbb{N}_r^{(l)})). \tag{6.9}$$

Определение 6.9. *Шифром замены с ограниченным ключом* назовем семейство:

$$\Sigma_O = (\Sigma_O^{(l)},\ l \in \mathbb{N};\ K,\ P(K);\ \psi_d).$$

При этом совокупность (6.9) будем называть l-ым опорным шифром шифра Σ_O.

6.7. Совершенные шифры замены

Для любого шифртекста $\overline{v} = v_1 \dots v_l \in V^{(l)}$ определим вероятность $P_{V^{(l)}}(\overline{v})$. Для шифра замены с неограниченным ключом Σ_H определим:

$$P_{V^{(l)}}(\overline{v}) = \sum_{\substack{(\overline{u},\overline{\jmath})\in U^{(l)}\times\mathbb{N}_r^l \\ E_{\overline{\jmath}}(\overline{u})=\overline{v}}} P_{U^{(l)}}(\overline{u}) \cdot P_{\mathbb{N}_r^l}(\overline{\jmath}).$$

Для шифра замены с ограниченным ключом Σ_O положим:

$$P_{V^{(l)}}(\overline{v}) = \sum_{\substack{(\overline{u},\overline{\jmath})\in U^{(l)}\times\mathbb{N}_r^{(l)} \\ E_{\overline{\jmath}}(\overline{u})=\overline{v}}} P_{U^{(l)}}(\overline{u}) \cdot P_{\mathbb{N}_r^{(l)}}(\overline{\jmath}).$$

Также определим условные вероятности:

$$P_{U^{(l)}|V^{(l)}}(\overline{u}|\overline{v}), \quad P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u})$$

для $\overline{u} = u_1 \dots u_l \in U^{(l)}$, $\overline{v} = v_1 \dots v_l \in V^{(l)}$.

Для шифра замены с неограниченным ключом Σ_H определим:

$$P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u}) = \sum_{\overline{\jmath}\in\mathbb{N}_r^l(\overline{u},\overline{v})} P_{\mathbb{N}_r^l}(\overline{\jmath}), \tag{6.10}$$

где:

$$\mathbb{N}_r^l(\overline{u},\overline{v}) = \{\overline{\jmath} \in \mathbb{N}_r^l \mid E_{\overline{\jmath}}(\overline{u}) = \overline{v}\},$$

а распределение $P(\mathbb{N}_r^l)$ является априорным распределением вероятностей на множестве ключевых потоков длины l.

Для шифра замены с ограниченным ключом Σ_O определим:

$$P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u}) = \sum_{\overline{\jmath}\in\mathbb{N}_r^{(l)}(\overline{u},\overline{v})} P_{\mathbb{N}_r^{(l)}}(\overline{\jmath}),$$

где:

$$\mathbb{N}_r^{(l)}(\overline{u},\overline{v}) = \{\overline{\jmath} \in \mathbb{N}_r^{(l)} \mid E_{\overline{\jmath}}(\overline{u}) = \overline{v}\},$$

а значение $P_{\mathbb{N}_r^{(l)}}(\overline{\jmath})$ определяется по формуле (6.8).

Вероятность $P_{U^{(l)}|V^{(l)}}(\overline{u}|\overline{v})$ вычисляется по формуле:

$$P_{U^{(l)}|V^{(l)}}(\overline{u}|\overline{v}) = \frac{P_{U^{(l)}}(\overline{u}) \cdot P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u})}{P_{V^{(l)}}(\overline{v})}.$$

В следующих утверждениях под шифром замены понимается либо шифр Σ_H, либо шифр Σ_O.

Определение 6.10. Шифр замены называется совершенным, если для любого натурального числа $l \in \mathbb{N}$ и любых $\overline{u} \in U^{(l)}$, $\overline{v} \in V^{(l)}$ выполнено равенство:

$$P_{U^{(l)}|V^{(l)}}(\overline{u}|\overline{v}) = P_{U^{(l)}}(\overline{u}).$$

Лемма 6.10. Для шифра замены следующие условия эквивалентны:

(i) для любого $l \in \mathbb{N}$ и любых $\overline{u} \in U^{(l)}$, $\overline{v} \in V^{(l)}$ выполнено равенство:

$$P_{U^{(l)}|V^{(l)}}(\overline{u}|\overline{v}) = P_{U^{(l)}}(\overline{u});$$

(ii) для любого $l \in \mathbb{N}$ и любых $\overline{u} \in U^{(l)}$, $\overline{v} \in V^{(l)}$ выполнено равенство:

$$P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u}) = P_{V^{(l)}}(\overline{v});$$

(iii) для любого $l \in \mathbb{N}$ и любых $\overline{u}_1, \overline{u}_2 \in U^{(l)}$, $\overline{v} \in V^{(l)}$ выполнено равенство:

$$P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u}_1) = P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{u}_2).$$

Лемма 6.11. Шифр замены является совершенным тогда и только тогда, когда для любого натурального l его l-й опорный шифр является совершенным шифром.

Лемма 6.12. Пусть шифр замены с неограниченным ключом Σ_H (с ограниченным ключом Σ_O) является совершенным. Тогда для данного шифра будут выполнены следующие свойства:

(i) для любого натурального l и любых $\overline{u} \in U^{(l)}$, $\overline{v} \in V^{(l)}$ найдется такой ключевой поток $\overline{\jmath} \in \mathbb{N}_r^l$ ($\overline{\jmath} \in \mathbb{N}_r^{(l)}$), что $E_{\overline{\jmath}}(\overline{u}) = \overline{v}$, т.е.:

$$|\mathbb{N}_r^l(\overline{u}, \overline{v})| \geqslant 1 \quad (|\mathbb{N}_r^{(l)}(\overline{u}, \overline{v})| \geqslant 1).$$

(ii) для любого натурального числа l справедливы следующие неравенства:

$$|U^{(l)}| \leqslant |V^{(l)}| \leqslant |\mathbb{N}_r^l|$$
$$(|U^{(l)}| \leqslant |V^{(l)}| \leqslant |\mathbb{N}_r^{(l)}|).$$

Теорема 6.9. Пусть последовательность $\{|U^{(l)}|\}_{l\geqslant 1}$ не является ограниченной:

$$\overline{\lim_{l\to\infty}}|U^{(l)}| = +\infty.$$

Тогда шифр замены с ограниченным ключом не является совершенным.

Доказательство. Как уже было отмечено ранее, для любого натурального числа l выполняется такое неравенство:

$$|\mathbb{N}_r^{(l)}| = \quad |\{\psi_d(k,l) \mid k \in K\}| \leqslant |K|.$$

Поскольку множество ключей K конечно, то последовательность $\{|\mathbb{N}_r^{(l)}|\}_{l\geqslant 1}$ является ограниченной. По условию теоремы последовательность $\{|U^{(l)}|\}_{l\geqslant 1}$ не является ограниченной, поэтому, в силу леммы 6.12, шифр замены с ограниченным ключом не является совершенным, так как в этом случае, как минимум, последовательность $\{|\mathbb{N}_r^{(l)}|\}_{l\geqslant 1}$ должна быть неограниченной. □

Теорема 6.10 (**достаточные условия совершенного шифра замены с неограниченным ключом**)**.** Пусть шифр замены Σ_H обладает следующими условиями:

(i) правила зашифрования E_1, E_2,. . . , E_r шифра Σ_H обладают тем свойством, что для любого $u \in U$ и любого $v \in V$ найдется, и притом единственный, элемент $j = j(u,v) \in \mathbb{N}_r$, что $E_j(u) = v$ (другими словами, для любого $u \in U$ и любого $v \in V$ выполнено равенство $|\mathbb{N}_r(u,v)| = 1$);

(ii) распределение вероятностей $P(\mathbb{N}_r)$ является равномерным, т.е. для любого $j \in \mathbb{N}_r$ выполнено равенство $P_{\mathbb{N}_r}(j) = \dfrac{1}{|\mathbb{N}_r|}$.

Тогда шифр Σ_H является совершенным, причем для любого $l \in \mathbb{N}$ выполнено равенство $|V^{(l)}| = r^l$ и распределение вероятностей $P(V^{(l)})$ будет являться равномерным.

Доказательство. Зафиксируем произвольное натуральное число l. Из пункта (i) следует, что для любых $\overline{u} = u_1 \dots u_l \in U^{(l)}$ и $\overline{v} = v_1 \dots v_l \in V^{(l)}$ найдется единственный ключевой поток

$\overline{j} = j_1 \dots j_l \in \mathbb{N}_r^l$, зависящий от $\overline{u}$ и $\overline{v}$, что:

$$E_{\overline{j}}(\overline{u}) = E_{j_1}(u_1) \dots E_{j_l}(u_l) = v_1 \dots v_l = \overline{v}.$$

Из данного свойства и того, что:

$$P_{\mathbb{N}_r^l}(\overline{j}) = P_{\mathbb{N}_r}(j_1) \cdot \dots \cdot P_{\mathbb{N}_r}(j_l) = \frac{1}{r^l}$$

для любого $\overline{j} \in \mathbb{N}_r^l$ (пункт (ii)) следует, что шифр $\Sigma_H^{(l)}$ является совершенным (теорема 6.1).

Таким образом, учитывая лемму 6.11, следует, что шифр Σ_H является совершенным. □

Пример 6.7. Пусть $X = \mathbb{Z}_3^*$, $Y = \mathbb{Z}_4^*$, т.е. открытыми и шифрованными текстами являются все конечные слова соответственно в алфавитах $\mathbb{Z}_3$ и $\mathbb{Z}_4$. Определим множество шифрвеличин и шифробозначений следующим образом: $U = \mathbb{Z}_3$, $V = \mathbb{Z}_4$. Пусть также имеется 4 простые замены $E = \{E_1, E_2, E_3, E_4\}$ и матрица зашифрования:

$\mathbb{N}_4 \diagdown U$	0	1	2
1	0	1	2
2	1	2	3
3	2	3	0
4	3	0	1

Пусть, например, требуется зашифровать открытый текст $\overline{u} = 20101 \in \mathbb{Z}_3^5$. Предположим, что случайный генератор ψ_c сгенерировал ключевую последовательность $\overline{j} = 31241 \in \mathbb{N}_4^5$. Тогда шифрованный текст будет иметь такой вид:

$$\overline{v} = E_{\overline{j}}(\overline{u}) = E_3(2)E_1(0)E_2(1)E_4(0)E_1(1) = 00231.$$

Заметим при этом, что если случайный генератор ψ_c вырабатывает равновероятные гаммы, то данный шифр является совершенным (теорема 6.10).

Теорема 6.11. Пусть для шифра Σ_H выполнено равенство: $|U| = |\mathbb{N}_r| = |V|$. Тогда шифр Σ_H является совершенным тогда и только тогда, когда выполнены следующие условия:

(i) правила зашифрования $E_1, E_2, \ldots, E_r$ шифра Σ_H обладают тем свойством, что для любых $u \in U$, $v \in V$ найдется, и притом единственный, элемент $j = j(u, v) \in \mathbb{N}_r$, что $E_j(u) = v$;

(ii) распределение вероятностей $P(\mathbb{N}_r)$ является равномерным.

Доказательство следует из теоремы Шеннона и теоремы 6.10. □

Теорему 6.11 можно переформулировать следующим образом.

Теорема 6.12. Пусть $|U| = |\mathbb{N}_r| = |V|$. Шифр Σ_H является совершенным тогда и только тогда, когда опорный для Σ_H шифр:

$$\Sigma = (U, \mathbb{N}_r, V, E, D, P(U), P(\mathbb{N}_r))$$

является совершенным.

Пусть для шифра Σ_H выполнено равенство $|U| = |\mathbb{N}_r| = |V|$ и распределение вероятностей $P(\mathbb{N}_r)$ является равномерным. Тогда из теоремы 6.11 следует, что шифр Σ_H является совершенным тогда и только тогда, когда матрица зашифрования опорного шифра для Σ_H :

$\mathbb{N}_r \diagdown U$	u_1	$\ldots$	u_r
1	$E_1(u_1)$	$\ldots$	$E_1(u_r)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
r	$E_r(u_1)$	$\ldots$	$E_r(u_r)$

является латинским квадратом.

Следствие 6.5 (теоремы 6.11)**.** Шифры табличного и модульного гаммирования с равновероятной гаммой являются совершенными шифрами замены.

Рассмотрим еще один критерий совершенных шифров замены с неограниченным ключом в классе шифров с равномерным распределением вероятностей на множестве $\mathbb{N}_r$.

Теорема 6.13. Шифр Σ_H с равномерным распределением вероятностей $P(\mathbb{N}_r)$ является совершенным тогда и только тогда, когда для любых $u_1, u_2 \in U$, $v \in V$ выполнено равенство:

$$|\mathbb{N}_r(u_1, v)| = |\mathbb{N}_r(u_2, v)|. \tag{6.11}$$

Доказательство Пусть шифр Σ_H является совершенным. Тогда, в частности, опорный шифр шифра Σ_H будет являться совершенным (лемма 6.11). Поэтому равенство (6.11) следует из теоремы 6.3.

Обратно, пусть выполнено равенство (6.11) для шифра Σ_H. Зафиксируем произвольное значение $l \in \mathbb{N}$. Заметим, что для любых $\overline{u} = u_1 \dots u_l \in U^{(l)}$ и $\overline{v} = v_1 \dots v_l \in V^{(l)}$ найдется такой ключевой поток $\overline{j} = j_1 \dots j_l \in \mathbb{N}_r^l$, зависящий от $\overline{u}$ и $\overline{v}$, что:

$$E_{\overline{j}}(\overline{u}) = E_{j_1}(u_1) \dots E_{j_l}(u_l) = v_1 \dots v_l = \overline{v}.$$

Далее, зафиксируем произвольным образом некоторые значения $\overline{a} = a_1 \dots a_l \in U^{(l)}$, $\overline{b} = b_1 \dots b_l \in U^{(l)}$, $\overline{v} = v_1 \dots v_l \in V^{(l)}$. Тогда:

$$|\mathbb{N}_r(\overline{a}, \overline{v})| = \prod_{i=1}^{l} |\mathbb{N}_r(a_i, v_i)| = \prod_{i=1}^{l} |\mathbb{N}_r(b_i, v_i)| = |\mathbb{N}_r(\overline{b}, \overline{v})|.$$

Таким образом, из теоремы 6.3 следует, что шифр $\Sigma_H^{(l)}$ является совершенным. В силу произвольности значения l следует совершенность шифра Σ_H. □

Следствие 6.6. Пусть для шифра Σ_H выполнено равенство $|V| = |\mathbb{N}_r|$ и распределение вероятностей $P(\mathbb{N}_r)$ является равномерным. Шифр Σ_H является совершенным тогда и только тогда, когда $|\mathbb{N}_r(u, v)| = 1$ для любых $u \in U$ и $v \in V$.

Рассмотрим задачу построения совершенного шифра Σ_H по заданному множеству шифрвеличин U и множеству $\mathbb{N}_r$ с распределением вероятностей $P(\mathbb{N}_r)$.

Теорема 6.14. Для заданных U, $|U| = n$, $\mathbb{N}_r$, $P(\mathbb{N}_r)$ существует совершенный шифр Σ_H тогда и только тогда, когда найдется такое натуральное число s и n разбиений множества $\mathbb{N}_r$:

$$\mathbb{N}_r = K_{11} \cup K_{12} \cup \dots \cup K_{1s}, \quad K_{1i} \cap K_{1j} = \varnothing, \quad 1 \leqslant i < j \leqslant s,$$

$$\mathbb{N}_r = K_{21} \cup K_{22} \cup \ldots \cup K_{2s}, \quad K_{2i} \cap K_{2j} = \varnothing, \quad 1 \leqslant i < j \leqslant s, \tag{6.12}$$

$$\ldots$$

$$\mathbb{N}_r = K_{n1} \cup K_{n2} \cup \ldots \cup K_{ns}, \quad K_{ni} \cap K_{nj} = \varnothing, \quad 1 \leqslant i < j \leqslant s,$$

для которых выполнены следующие условия:

1) $K_{it} \cap K_{jt} = \varnothing$, $1 \leqslant i < j \leqslant n$, $t = 1, \ldots, s$;

2) для любых $1 \leqslant i < j \leqslant n$, $t = 1, \ldots, s$ выполнено равенство:

$$\sum_{k \in K_{it}} P_{\mathbb{N}_r}(k) = \sum_{k \in K_{jt}} P_{\mathbb{N}_r}(k).$$

Доказательство. *Необходимое* условие следует из теоремы 6.4.

Достаточность. Пусть для U, $\mathbb{N}_r$, $P(\mathbb{N}_r)$, найдется такое s и n таких разбиений (6.12), для которых выполнены условия 1, 2. Пусть V — некоторое множество шифробозначений, $|V| = s$. Составим матрицу зашифрования над элементами множества V для опорного шифра Σ также, как и в теореме 3. Зафиксируем некоторое натуральное l. Пусть $\overline{a} = a_1 \ldots a_l \in U^{(l)}$, $\overline{b} = b_1 \ldots b_l \in U^{(l)}$, $\overline{v} = v_1 \ldots v_l \in V^{(l)}$. Тогда:

$$P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{a}) = \prod_{i=1}^{l} P_{V|U}(v_i|a_i) = \prod_{i=1}^{l} P_{V|U}(v_i|b_i) = P_{V^{(l)}|U^{(l)}}(\overline{v}|\overline{b}),$$

где второе равенство следует из теоремы 6.4. Поэтому из леммы 6.10 следует, что шифр $\Sigma_H^{(l)}$ является совершенным. □

Следствие 6.7. Пусть для заданных U, $\mathbb{N}_r$, $P(\mathbb{N}_r)$ существует совершенный шифр. Тогда для любого множества шифрвеличин $\widetilde{U}$, $|\widetilde{U}| \leqslant |U|$, и для заданных $\mathbb{N}_r$, $P(\mathbb{N}_r)$ существует совершенный шифр Σ_H.

Следствие 6.8. Для заданных U, $|U| = n$, $\mathbb{N}_r$, $P(\mathbb{N}_r)$, V, $|V| = s$, существует совершенный шифр Σ_H тогда и только тогда, когда найдется n таких разбиений (6.12), для которых выполнены условия 1, 2 предыдущей теоремы.

Следствие 6.9. Для заданных V, $|V| = s$, $\mathbb{N}_r$, $P(\mathbb{N}_r)$, существует совершенный шифр Σ_H тогда и только тогда, когда

найдется такое n и такие разбиения (6.12), для которых выполнены условия 1, 2 предыдущей теоремы.

6.8. $(k|y)$-совершенные шифры Σ_H

Далее везде предполагается, что для любого натурального l выполнены равенства $U^{(l)} = U^l$, $V^{(l)} = V^l$.

Пусть $\overline{\jmath} \in \mathbb{N}_r^l$, $\overline{v} \in V^l$. Обозначим:

$$U^l(\overline{\jmath}, \overline{v}) = \{\overline{u} \in U^l \mid E_{\overline{\jmath}}(\overline{u}) = \overline{v}\}.$$

Заметим, что множество $U^l(\overline{\jmath}, \overline{v})$ либо пусто, либо одноэлементно.

Определим условные вероятности $P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath})$ и $P_{\mathbb{N}_r^l|V^l}(\overline{\jmath}|\overline{v})$:

$$P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath}) = \begin{cases} P_{U^l}(\overline{u}), & U^l(\overline{\jmath}, \overline{v}) = \{\overline{u}\}, \\ 0, & U^l(\overline{\jmath}, \overline{v}) = \varnothing, \end{cases}$$

$$P_{\mathbb{N}_r^l|V^l}(\overline{\jmath}|\overline{v}) = \frac{P_{\mathbb{N}_r^l}(\overline{\jmath}) \cdot P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath})}{P_{V^l}(\overline{v})}.$$

Говорят, что шифр Σ_H является $(k|y)$-совершенным, если для любого натурального l шифр Σ_H^l является $(k|y)$-совершенным.

Лемма 6.13. Для шифра Σ_H следующие условия эквивалентны:

(i) для любого $l \in \mathbb{N}$ и любых $\overline{\jmath} \in \mathbb{N}_r^l$, $\overline{v} \in V^l$ выполнено равенство $P_{\mathbb{N}_r^l|V^l}(\overline{\jmath}|\overline{v}) = P_{\mathbb{N}_r^l}(\overline{\jmath})$;

(ii) для любого $l \in \mathbb{N}$ и любых $\overline{\jmath} \in \mathbb{N}_r^l$, $\overline{v} \in V^l$ выполнено равенство $P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath}) = P_{V^l}(\overline{v})$;

(iii) для любого $l \in \mathbb{N}$ и любых $\overline{\jmath}_1, \overline{\jmath}_2 \in \mathbb{N}_r^l$, $\overline{v} \in V^l$ выполнено равенство $P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath}_1) = P_{V^l|\mathbb{N}_r^l}(\overline{v}|\overline{\jmath}_2)$.

Теорема 6.15. Пусть для шифра Σ_H выполнены следующие условия:

(i) $|U| = |V|$;

(ii) для любого $l \in \mathbb{N}$ распределение вероятностей $P(U^l)$ является равномерным.

Тогда шифр Σ_H является $(k|y)$-совершенным.

Доказательство. Зафиксируем $l \in \mathbb{N}$. Пусть $j_1 \dots j_l \in \mathbb{N}_r^l$, $v_1 \dots v_l \in V^l$. Тогда найдется, и притом единственный, открытый текст $u_1 \dots u_l \in U^l$, такой что $E_{j_1 \dots j_l}(u_1 \dots u_l) = v_1 \dots v_l$. Поэтому:

$$P_{V^l|\mathbb{N}_r^l}(v_1 \dots v_l | j_1 \dots j_l) = P_{U^l}(u_1 \dots u_l) = \frac{1}{|U|^l} = \frac{1}{|V|^l}.$$

Поэтому для любых $\overline{j}_1, \overline{j}_2 \in \mathbb{N}_r^l$, $\overline{v}_1, \overline{v}_2 \in V^l$ выполнены равенства:

$$P_{V^l|\mathbb{N}_r^l}(\overline{v}_1 | \overline{j}_1) = \frac{1}{|V|^l} = P_{V^l|\mathbb{N}_r^l}(\overline{v}_2 | \overline{j}_2).$$

Таким образом, из леммы 6.13 следует, что шифр Σ_H^l является $(k|y)$-совершенным. □

Пусть распределение вероятностей $P(U)$ индуцирует распределение вероятностей $P(U^l)$ следующим образом:

$$P_{U^l}(u_1 \dots u_l) = \prod_{i=1}^{l} P_U(u_i), \quad u_1 \dots u_l \in U^l, \quad l \in \mathbb{N}. \tag{6.13}$$

Лемма 6.14. Для заданных множеств U и V существует $(k|y)$-совершенный шифр Σ_H с распределением вероятностей (6.13) тогда и только тогда, когда $|U| = |V|$.

Доказательство. *Достаточность.* Пусть для произвольного фиксированного $m > 1$ найдутся такие перестановки $\sigma_1, \dots, \sigma_m \in S_n$, где $n = |U| = |V|$, для которых выполнены условия:

$$P_U(u_{\sigma_i(s)}) = P_U(u_{\sigma_j(s)}), \quad 1 \leqslant i < j \leqslant m, \quad s = 1, \dots, n.$$

Составим матрицу зашифрования размера $m \times n$ для опорного шифра следующим образом: на позицию $(i, \sigma_i(s))$ поставим v_s, $i = 1, \dots, m$, $s = 1, \dots, n$. Так как $\mathbb{N}_m(i, v_s) = \{u_{\sigma_i(s)}\}$, то для любого $l \in \mathbb{N}$ и любых $v_{i_1} \dots v_{i_l} \in V^l$, $a_1 \dots a_l \in \mathbb{N}_m^l$, $b_1 \dots b_l \in \mathbb{N}_m^l$ выполнены равенства:

$$P_{V^l|\mathbb{N}_m^l}(v_{i_1} \dots v_{i_l} | a_1 \dots a_l) = P_{U^l}(u_{\sigma_{a_1}(i_1)} \dots u_{\sigma_{a_l}(i_l)}) =$$

$$= \prod_{t=1}^{l} P_U(u_{\sigma_{a_t}(i_t)}) = \prod_{t=1}^{l} P_U(u_{\sigma_{b_t}(i_t)}) =$$

$$= P_{U^l}(u_{\sigma_{b_1}(i_1)} \dots u_{\sigma_{b_l}(i_l)}) = P_{V^l|\mathbb{N}_m^l}(v_{i_1} \dots v_{i_l}|b_1 \dots b_l).$$

Таким образом, из леммы 6.13 следует, шифр Σ_H^l является $(k|y)$-совершенным.

Необходимое условие следует из леммы 6.9. □

Теорема 6.16. Для заданных множеств $V = \{v_1, \dots, v_n\}$, $U = \{u_1, \dots, u_n\}$ с равномерным распределением $P(U^l)$ для любого $l \in \mathbb{N}$, $\mathbb{N}_r$ с распределением вероятностей $P_{\mathbb{N}_r}$ существует одновременно совершенный и $(k|y)$-совершенный шифр Σ_H тогда и только тогда, когда найдется такая матрица $A = A(\mathbb{N}_r)$ порядка $n \times n$, каждый элемент которой является непустым подмножеством в $\mathbb{N}_r$, для которой выполнены следующие условия:

1) каждая строка и каждый столбец матрицы A является разбиением множества $\mathbb{N}_r$ на непересекающиеся подмножества;

2) для любых $i = 1, \dots, n$, $j = 1, \dots, n$ выполнено равенство:

$$\sum_{k \in A_{ij}} P_{\mathbb{N}_r}(k) = \frac{1}{n}.$$

Доказательство. *Необходимое* условие следует из теоремы 6.8.

Достаточность. Составим матрицу зашифрования над элементами множества V для опорного шифра Σ также, как и в теореме 6.8. Зафиксируем некоторое натуральное l. Пусть $\overline{a} = a_1 \dots a_l \in U^l$, $\overline{b} = b_1 \dots b_l \in U^l$, $\overline{v} = v_1 \dots v_l \in V^l$. Тогда:

$$P_{V^l|U^l}(\overline{v}|\overline{a}) = \prod_{i=1}^{l} P_{V|U}(v_i|a_i) = \prod_{i=1}^{l} P_{V|U}(v_i|b_i) = P_{V^l|U^l}(\overline{v}|\overline{b}),$$

где второе равенство следует из теоремы 6.8. Поэтому из леммы 6.10 и теоремы 6.5 следует, что шифр Σ_H^l является совершенным и $(k|y)$-совершенным. □

6.9. Вопросы имитостойкости шифров

Заметим, что потенциальный противник может осуществлять не только пассивные действия относительно передаваемых по каналу связи сообщений, которые заключаются, например в

подслушивании или перехвате сообщений, но также и активные атаки, заключающиеся в *имитации* или *подмене* сообщения.

Если канал связи готов к работе и на приеме установлены действующие ключи, но в данный момент времени никакого сообщения не передается, то в этом случае противником может быть предпринята попытка имитации сообщения. В таком случае противник может выбрать некоторый элемент $y \in Y$ и послать его от имени законного отправителя. При этом он будет рассчитывать на то, что на действующем ключе его криптограмма при расшифровании будет воспринята как некий осмысленный открытый текст. Чем больше вероятность этого события, тем успешнее будет попытка имитации.

Рассмотрим вероятностное пространство $(\Omega = K, F_K, P_K)$. Зафиксируем $y \in Y$. Обозначим через $K(y)$ следующее множество:

$$K(y) = \{k \in K \mid y \in E_k(X)\}.$$

Под обозначением $K(y)$ будем также понимать событие из алгебры событий F_K, заключающееся в том, что при случайном выборе ключа $k \in K$ шифртекст y можно расшифровать на ключе k, т.е. $y \in E_k(X)$. Тогда событию $K(y)$ будут благоприятствовать все элементы из множества $K(y)$ и только они. Поэтому:

$$P(K(y)) = \sum_{k \in K(y)} P_K(k).$$

Поскольку противник имеет возможность выбора $y \in Y$, его шансы на успех при имитации сообщения выражаются такой величиной:

$$P_{im} = \max_{y \in Y} P(K(y)).$$

Если же в данный момент передается некоторое сообщение $y \in Y$ (которое получено из открытого текста $x \in X$ на ключе $k \in K$), то противник может заменить его на $\widetilde{y} \in Y$, отличный от y. При этом он будет рассчитывать на то, что на действующем ключе k криптограмма $\widetilde{y}$ будет воспринята как некий осмысленный открытый текст $\widetilde{x}$, отличный от x. Чем больше

вероятность этого события, тем успешнее будет попытка подмены. Пусть "$K(\widetilde{y}) \mid K(y)$" — событие, заключающееся в попытке подмены сообщения y сообщением $\widetilde{y}$. Применяя теорему о произведении вероятностей, получаем, что:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{P(K(y) \cap K(\widetilde{y}))}{P(K(y))} = \\ = \frac{\sum_{k \in K(y,\widetilde{y})} P_K(k)}{\sum_{k \in K(y)} P_K(k)},$$

где $K(y, \widetilde{y}) = K(y) \cap K(\widetilde{y})$. Тогда вероятность подмены сообщения будет вычисляться по следующей формуле:

$$P_{podm} = \max_{\substack{y,\widetilde{y} \in Y \\ y \neq \widetilde{y}}} P(K(\widetilde{y}) \mid K(y)).$$

Теорема 6.17. Для любого шифра Σ_B справедливо неравенство:

$$P_{im} \geqslant \frac{|X|}{|Y|}. \tag{6.14}$$

Равенство в (6.14) имеет место тогда и только тогда, когда для любого $y \in Y$ выполнено равенство:

$$P(K(y)) = \frac{|X|}{|Y|}.$$

Доказательство. 1. Рассмотрим сначала случай, когда шифр Σ_B имеет равновероятное распределение ключей. В этом случае:

$$P(K(y)) = \frac{|K(y)|}{|K|}.$$

Поэтому:

$$P_{im} = \max_{y \in Y} \frac{|K(y)|}{|K|}. \tag{6.15}$$

Заметим, что справедливо следующее равенство:

$$\sum_{y \in Y} |K(y)| = |X| \cdot |K|. \tag{6.16}$$

Действительно, рассмотрим матрицу зашифрования для шифра Σ_B, строки которой занумерованы ключами, а столбцы —

открытыми текстами. Элементами данной матрицы являются элементы множества Y, причем на пересечении строки с номером $k \in K$ и столбца с номером $x \in X$ расположен элемент $y = E_k(x)$. В силу определения понятия шифра, никакая строка данной матрицы не содержит двух одинаковых элементов из множества Y. Поэтому если некоторый $y_0 \in Y$ в матрице зашифрования встречается n_0 раз, то (с учетом того, что y_0 расположен в различных n_0 строках), данный шифрованный текст можно расшифровать с помощью n_0 различных ключей, т.е. $|K(y_0)| = n_0$. Поэтому число $|K(y)|$ можно интерпретировать как количество появлений элемента $y \in Y$ в матрице зашифрования. Так как данная матрица имеет размерность $|K| \times |X|$, то это показывает справедливость формулы (6.16).

Так как:

$$\sum_{y \in Y} |K(y)| \leqslant \sum_{y \in Y} \max_{y \in Y} |K(y)| = \max_{y \in Y} |K(y)| \cdot |Y|,$$

то из (6.16) следует такое неравенство:

$$\max_{y \in Y} |K(y)| \geqslant \frac{|X| \cdot |K|}{|Y|}.$$

Из данного неравенства и из (6.15) следует неравенство (6.14).

2. Рассмотрим более общий случай. Пусть теперь шифр Σ_B произвольный. Заметим, что справедливо такое равенство:

$$\sum_{y \in Y} P(K(y)) = \sum_{y \in Y} \sum_{k \in K(y)} P_K(k) = \sum_{k \in K} P_K(k) \cdot |X| = |X|. \quad (6.17)$$

Действительно, рассуждая аналогично, как и при доказательстве равенства (6.16), заметим, что в данной двойной сумме каждый ключ учитывается ровно $|X|$ раз.

Учитывая равенство (6.17), получаем:

$$|X| = \sum_{y \in Y} P(K(y)) \leqslant \sum_{y \in Y} \max_{y \in Y} P(K(y)) = \sum_{y \in Y} P_{im} = |Y| \cdot P_{im},$$

откуда немедленно вытекает неравенство (6.14).

Покажем вторую часть теоремы. Если для любого $y \in Y$ выполнено равенство $P(K(y)) = \dfrac{|X|}{|Y|}$, то очевидно, что $P_{im} = \dfrac{|X|}{|Y|}$.

Обратно, предположим, что $P_{im} = \dfrac{|X|}{|Y|}$. Предположим, что найдется такой $y_0 \in Y$, что:

$$P(K(y_0)) < \frac{|X|}{|Y|} = P_{im}.$$

Тогда, учитывая равенство (6.17), получаем:

$$|X| = \sum_{y \in Y} P(K(y)) = P(K(y_0)) + \sum_{y \in Y \setminus \{y_0\}} P(K(y)) <$$

$$< P_{im} + (|Y| - 1) \cdot P_{im} = |Y| \cdot P_{im}.$$

Поэтому $P_{im} > \dfrac{|X|}{|Y|}$. Противоречие.

Следовательно, $P(K(y)) = \dfrac{|X|}{|Y|}$ для любого $y \in Y$. □

Пример 6.8. Приведем примеры шифров, у которых для P_{im} достигается нижняя граница: $P_{im} = \dfrac{|X|}{|Y|}$.

1. Пусть для шифра Σ_B выполнено условие $|X| = |Y|$. Тогда все строки матрицы зашифрования данного шифра являются перестановками элементов множества Y. Поэтому для любого $y \in Y$ будет выполнено равенство $P(K(y)) = 1 = \dfrac{|X|}{|Y|}$. Следовательно, $P_{im} = 1 = \dfrac{|X|}{|Y|}$.

Такие шифры максимально уязвимы к угрозам имитации сообщения. Теорема 6.17 показывает, что имитостойкость растет пропорционально отношению $\dfrac{|Y|}{|X|}$. Поэтому для имитозащиты вводят избыточность в передаваемое сообщение.

2. Рассмотрим шифр Σ_B с такой матрицей зашифрования:

$K \diagdown X$	x_1	$\ldots$	x_N
k_1	y_{11}	$\ldots$	y_{1N}
$\ldots$	$\ldots$	$\ldots$	$\ldots$
k_M	y_{M1}	$\ldots$	y_{MN}

где все элементы y_{ij} данной матрицы различные. Данный шифр определяется условием $E_{k_i}(x_s) = E_{k_j}(x_t)$ тогда и только тогда, когда $k_i = k_j$ и $x_s = x_t$. Поэтому $|Y| = |X| \cdot |K|$. Пусть распределение вероятностей на множестве K является равномерным. Тогда для любого $y \in Y$ будет выполняться равенство $P(K(y)) = \dfrac{|1|}{|K|} = \dfrac{|X|}{|Y|}$, откуда по теореме 6.17 следует, что $P_{im} = \dfrac{|X|}{|Y|}$.

Теорема 6.18. Для любого шифра Σ_B справедливо неравенство:

$$P_{podm} \geqslant \frac{|X| - 1}{|Y| - 1}. \tag{6.18}$$

Равенство в (6.18) имеет место тогда и только тогда, когда для любых $y, \widetilde{y} \in Y$, $y \neq \widetilde{y}$, выполнено равенство:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{|X| - 1}{|Y| - 1}.$$

Доказательство. 1. Как и в предыдущей теореме рассмотрим сначала случай, когда шифр Σ_B имеет равновероятное распределение ключей. В этом случае:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{|K(y, \widetilde{y})|}{|K(y)|}.$$

Поэтому:

$$P_{podm} = \max_{\substack{y, \widetilde{y} \in Y \\ y \neq \widetilde{y}}} \frac{|K(y, \widetilde{y})|}{|K(y)|}.$$

Зафиксируем некоторый элемент $y \in Y$. Заметим, что справедливо такое равенство:

$$\sum_{\widetilde{y} \in Y \setminus \{y\}} |K(y, \widetilde{y})| = (|X| - 1) \cdot |K(y)|. \tag{6.19}$$

Действительно, рассмотрим, как и в предыдущей теореме, матрицу зашифрования. Вычеркнем в ней все те строки, которые не содержат шифрованного сообщения y. Останется ровно $|K(y)|$ строк. В полученной матрице $(|X|-1) \cdot |K(y)|$ элементов из множества Y будут отличны от y. Если некоторый элемент $y_0 \in Y$

в полученной матрице (размера $(|X|-1)\cdot|K(y)|$) встречается n_0 раз, то (с учетом того, что y_0 расположен в различных n_0 строках), данный элемент можно расшифровать с помощью n_0 различных ключей. Поэтому $|K(y,y_0)|=n_0$. Следовательно, $|K(y,y_0)|$ — количество появлений элемента $y_0\in Y$ в полученной матрице.

Далее, поскольку количество слагаемых в сумме (6.19) равно $|Y|-1$, то справедливо следующее неравенство:

$$\max_{\widetilde{y}\in Y\setminus\{y\}}\frac{|K(y,\widetilde{y})|}{|K(y)|}\geqslant\frac{|X|-1}{|Y|-1}.$$

Учитывая данное неравенство и то, что y — произвольный элемент из Y, имеем:

$$P_{podm}=\max_{\substack{y,\widetilde{y}\in Y\\ y\neq\widetilde{y}}}\frac{|K(y,\widetilde{y})|}{|K(y)|}\geqslant\max_{\widetilde{y}\in Y\setminus\{y\}}\frac{|K(y,\widetilde{y})|}{|K(y)|}\geqslant\frac{|X|-1}{|Y|-1}.$$

2. Перейдем к рассмотрению более общего случая. Пусть теперь шифр Σ_B является произвольным. Зафиксируем некоторый элемент $y\in Y$. Имеем такую цепочку равенств:

$$\sum_{\widetilde{y}\in Y\setminus\{y\}}P(K(\widetilde{y})\mid K(y))=$$

$$=\frac{1}{P(K(y))}\cdot\sum_{\widetilde{y}\in Y\setminus\{y\}}P(K(y)\ \cap\ K(\widetilde{y}))=$$

$$=\frac{1}{P(K(y))}\cdot\sum_{\widetilde{y}\in Y\setminus\{y\}}\ \sum_{k\in K(y,\widetilde{y})}P_K(k)=$$

$$=\frac{1}{P(K(y))}\cdot\left((|X|-1)\cdot\sum_{k\in K(y)}P_K(k)\right)=|X|-1.$$

Таким образом, получаем равенство:

$$\sum_{\widetilde{y}\in Y\setminus\{y\}}P(K(\widetilde{y})\mid K(y))=|X|-1. \tag{6.20}$$

Следовательно:

$$\max_{\widetilde{y}\in Y\setminus\{y\}}P(K(\widetilde{y})\mid K(y))\geqslant\frac{|X|-1}{|Y|-1}.$$

В силу произвольности $y \in Y$, из определения вероятности подмены сообщения немедленно вытекает равенство (6.18).

Покажем вторую часть теоремы. Если для любых $y, \widetilde{y} \in Y$, $y \neq \widetilde{y}$, выполнено равенство:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{|X|-1}{|Y|-1},$$

то очевидно, что $P_{podm} = \dfrac{|X|-1}{|Y|-1}$.

Обратно, пусть $P_{podm} = \dfrac{|X|-1}{|Y|-1}$. Предположим, что найдутся такие $y_1, y_2 \in Y$, $y_1 \neq y_2$, для которых выполнено строгое неравенство:

$$P(K(y_1) \mid K(y_2)) < \frac{|X|-1}{|Y|-1} = P_{podm}.$$

Тогда, с учетом равенства 6.20, получаем следующее:

$$|X| - 1 = \sum_{y \in Y \setminus \{y_1\}} P(K(y) \mid K(y_1)) =$$

$$= P(K(y_2) \mid K(y_1)) + \sum_{y \in Y \setminus \{y_1, y_2\}} P(K(y) \mid K(y_1)) <$$

$$< P_{podm} + (|Y| - 2) \cdot P_{podm} = (|Y| - 1) \cdot P_{podm}.$$

Поэтому $P_{podm} > \dfrac{|X|-1}{|Y|-1}$, что противоречит условию $P_{podm} = \dfrac{|X|-1}{|Y|-1}$. □

Пример 6.9. Рассмотрим примеры шифров, для которых выполнено равенство $P_{podm} = \dfrac{|X|-1}{|Y|-1}$.

1. Пусть $|X| = |Y|$. Тогда для любых $\widetilde{y}, y \in Y$:

$$P(K(\widetilde{y}) \mid K(y)) = 1 = \frac{|X|-1}{|Y|-1}.$$

Поэтому:

$$P_{podm} = 1 = \frac{|X|-1}{|Y|-1}.$$

2. Пусть $X = \{x_1, x_2\}$, $Y = \{y_1, y_2, y_3\}$, $K = \{k_1, k_2, k_3\}$ и матрица зашифрования имеет вид:

$K \diagdown X$	x_1	x_2
k_1	y_1	y_2
k_2	y_2	y_3
k_3	y_3	y_1

Пусть также распределение вероятностей на множестве K является равномерным. Тогда для любых $\widetilde{y}, y \in Y$:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{|K(\widetilde{y}, y)|}{|K(y)|} = \frac{1}{2} = \frac{|X| - 1}{|Y| - 1}.$$

Поэтому:

$$P_{podm} = \frac{1}{2} = \frac{|X| - 1}{|Y| - 1}.$$

3. Обобщим предыдущий случай. Пусть $|X| = m$, $|Y| = n$, $|K| = C_n^m$ и все строки матрицы зашифрования являются сочетаниями из n элементов множества Y по m. Пусть также распределение вероятностей $P(K)$ равномерно. Тогда:

$$P(K(\widetilde{y}) \mid K(y)) = \frac{|K(\widetilde{y}, y)|}{|K(y)|} = \frac{C_{n-2}^{m-2}}{C_{n-1}^{m-1}} = \frac{m-1}{n-1} = \frac{|X| - 1}{|Y| - 1}.$$

Следовательно:

$$P_{podm} = \frac{m-1}{n-1} = \frac{|X| - 1}{|Y| - 1}.$$

Заметим также, что:

$$P(K(y)) = \frac{|K(y)|}{|K|} = \frac{C_{n-1}^{m-1}}{C_n^m} = \frac{m}{n} = \frac{|X|}{|Y|}.$$

Поэтому:

$$P_{im} = \frac{m}{n} = \frac{|X|}{|Y|}.$$

6.10. Совершенные имитостойкие шифры

Предложение 6.1. Пусть $A = A(m, n)$ — некоторая матрица над множеством $Y = \{y_1, \ldots, y_m\}$, $m \geqslant n$, которая является латинским прямоугольником. Тогда если распределение

вероятностей $P(K)$ является равномерным, то для шифра Σ_B с матрицей зашифрования A выполнено равенство $P_{im} = n/m$.

Доказательство. Заметим, что из равномерости распределения $P(K)$ следует, что для любого $y \in Y$ выполнено равенство:

$$P(K(y)) = \frac{|K(y)|}{|K|}.$$

Так как произвольный элемент $y \in Y$ встречается ровно в n строках матрицы A, то $|K(y)| = n$ и, в силу произвольности y, из теоремы 6.17 следует, что $P_{im} = n/m$. □

Предложение 6.2. Пусть $A = A(n+1, n)$ — некоторая матрица над множеством $Y = \{y_1, \ldots, y_{n+1}\}$, которая является латинским прямоугольником. Тогда если распределение вероятностей $P(K)$ является равномерным, то для шифра Σ_B с матрицей зашифрования A выполнено равенство $P_{podm} = (n-1)/n$.

Доказательство. Заметим, что из равномерности распределения $P(K)$ следует, что для любых $\widetilde{y}, y \in Y$, $\widetilde{y} \neq y$, выполнено равенство:

$$P\big(K(\widetilde{y}) \mid K(y)\big) = \frac{|K(\widetilde{y}, y)|}{|K(y)|}.$$

Дополним матрицу A до латинского квадрата B размера $(n+1) \times (n+1)$ (см. [45]). Зафиксируем произвольный элемент $y_0 \in Y$. Так как матрица B является латинским квадратом, то элемент y_0 присутствует в последнем столбце матрицы B. Пусть он находится на позиции $(i_0, n+1)$. Это означает, что в матрице A элемент y_0 встречается во всех строках, кроме строки с номером i_0, а в i_0-й строке матрицы A расположены все элементы множества $Y \backslash \{y_0\}$. Поэтому для любого $y \in Y \backslash \{y_0\}$ выполнено равенство $|K(y, y_0)| = n - 1$. Также очевидно, что для любого $y \in Y$ выполнено равенство $|K(y)| = n$. В силу произвольности y_0, из теоремы 6.18 следует, что $P_{podm} = (n-1)/n$. □

Несложно проверить также следующее предложение.

Предложение 6.3. Пусть $B = B(m, m)$ — квадрат Виженера над множеством $Y = \{y_1, \ldots, y_m\}$. Составим из первых n

столбцов матрицы B матрицу A, где $1 \leqslant n \leqslant m-1$. Пусть $|K| = m$, $|X| = n$, матрица A является матрицей зашифрования для шифра Σ_B и распределение вероятностей $P(K)$ является равномерным. Тогда для шифра Σ_B выполнено равенство $P_{podm} = (n-1)/n$.

Пусть, как и ранее, Σ_H — шифр замены с неограниченным ключом, $\Sigma_H^{(l)}$ — l-я степень опорного шифра замены с неограниченным ключом.

В данном параграфе будем предполагать, что для любого натурального l выполнены равенства $U^{(l)} = U^l$, $V^{(l)} = V^l$. Обозначим через P_{im}^l вероятность имитации сообщения для шифра $\Sigma_H^{(l)}$, а через $P_{podm}^l(s)$ — вероятность подмены в сообщении длины l ровно s символов для шифра $\Sigma_H^{(l)}$, где $s \leqslant l$. Из теорем 6.17 и 6.18 следует, что если для некоторого шифра Σ_H выполнено равенство $|U| = |V|$, где U, V — множества шифрвеличин и шифробозначений соответственно, то $P_{im}^l = P_{podm}^l(s) = 1$ для любого натурального l и любого $s \leqslant l$, т.е. такие шифры максимально уязвимы к угрозам имитации и подмены сообщения.

Предложение 6.4. Пусть $A = A(n+1, n)$ — некоторая матрица над множеством шифробозначений $V = \{v_1, \ldots, v_{n+1}\}$, которая является латинским прямоугольником, и пусть матрица A является матрицей зашифрования для опорного шифра замены с неограниченным ключом Σ_H. Пусть также случайный генератор ключевых последовательностей φ_c из конструкции шифра Σ_H имеет равномерное распределение. Тогда для любого натурального l шифр $\Sigma_H^{(l)}$ является совершенным и выполнены следующие равенства:

$$P_{im}^l = \left(\frac{n}{n+1}\right)^l, \quad P_{podm}^l(s) = \left(\frac{n-1}{n}\right)^s.$$

Доказательство следует из теоремы 6.10 и предложений 6.1 и 6.2.

Пусть S_n — симметрическая группа степени n, $T^j \in S_n$ — циклическая перестановка на j позиций влево. Обозначим через $A_j = A_j(n, 2)$ матрицу размера $n \times 2$ над множеством $\mathbb{N}_n$,

имеющую такой вид:

$$A_j = \begin{pmatrix} 1 & 2 & \dots & n \\ T^j(1) & T^j(2) & \dots & T^j(n) \end{pmatrix}^T, \quad j = 1, \dots, n-1.$$

Из матриц A_j, $j = 1, \dots, n-1$, составим матрицу M размера $(n^2-n)\times 2$ путем последовательной графической записи матриц $A_1, \dots, A_{n-1}$ одной под другой.

Предложение 6.5. Пусть $M = M(n^2-n, 2)$ — матрица над множеством $V = \{v_1, \dots, v_n\}$, построенная выше, $r = n^2 - n$, $|U| = 2$ и пусть матрица M является матрицей зашифрования для опорного шифра замены с неограниченным ключом Σ_H. Пусть также случайный генератор ключевых последовательностей ψ_c из конструкции шифра Σ_H имеет равномерное распределение. Тогда для любого натурального l шифр Σ_H^l является совершенным и выполнены следующие равенства:

$$P_{im}^l = \left(\frac{2}{n}\right)^l, \quad P_{podm}^l(s) = \left(\frac{1}{n-1}\right)^s.$$

Доказательство следует из теорем 6.13, 6.17, 6.18.

Заметим, что в предложениях 6.4 и 6.5 $P_{im}^l \to 0$ при $l \to \infty$, $P_{podm}^l(s) \to 0$ при $s \to \infty$.

Пример 6.10 (пример имитостойкого совершенного шифра замены с неограниченным ключом)**.** Пусть исходный алфавит A совпадает со множеством шифрвеличин U, а алфавит B совпадает со множеством шифробозначений V, причем:

$$U = A = \mathbb{Z}_n, \quad V = B = \mathbb{Z}_{n+1}$$

для некоторого n. Пусть множества $E = \{E_0, E_1, \dots, E_n\}$ и $D = \{D_0, D_1, \dots, D_n\}$ состоят соответственно из правил зашифрования и расшифрования:

$$E_j : \mathbb{Z}_n \to \mathbb{Z}_{n+1}, \quad D_j : E_j(\mathbb{Z}_n) \to \mathbb{Z}_n, \quad j = 0, 1, \dots, n,$$

определенных по такому правилу:

$$E_j(x) = x + j \pmod{(n+1)}, \ x \in \mathbb{Z}_n,$$

$$D_j(y) = y - j \pmod{(n+1)},\ y \in \mathbb{Z}_{n+1}.$$

Опорный шифр будет иметь такой вид:

$$\Sigma = (\mathbb{Z}_n, \mathbb{Z}_{n+1}, \mathbb{Z}_{n+1}, E, D),$$

а матрица зашифрования опорного шифра примет следующий вид:

$\mathbb{Z}_{n+1} \diagdown \mathbb{Z}_n$	0	1	2	...	$n-2$	$n-1$
0	0	1	2	...	$n-2$	$n-1$
1	1	2	3	...	$n-1$	n
...	...	...	...	...	...	...
n	n	0	1	...	$n-3$	$n-2$

Заметим, что данная матрица зашифрования получается из матрицы всех циклических сдвигов влево множества $\mathbb{Z}_{n+1}$ путем вычеркивания последнего столбца.

Будем также считать, что выбор того или иного правила зашифрования из множества E происходит по равновероятному закону. Исходя из этого и определения правил зашифрования из множества E будет следовать, что опорный шифр Σ является совершенным (теорема 6.1).

Пусть l-я степень опорного шифра замены с неограниченным ключом имеет такой вид:

$$\Sigma_H^{(l)} = (\mathbb{Z}_n^l, \mathbb{Z}_{n+1}^l, \mathbb{Z}_{n+1}^l, E^{(l)}, D^{(l)}).$$

Будем считать, что имеется случайный генератор, который вырабатывает ключевые последовательности элементов из множества $\mathbb{Z}_{n+1}$ произвольной длины. Из теоремы 6.10 будет следовать, что шифр $\Sigma_H^{(l)}$ будет совершенным для любого l.

Вычислим вероятности имитации и подмены для опорного шифра Σ :

$$P_{im} = \frac{n}{n+1} = \frac{|U|}{|V|},$$

$$P_{podm} = \frac{n-1}{n} = \frac{|U|-1}{|V|-1}.$$

Для шифра $\Sigma_H^{(l)}$ выполнены следующие равенства:

$$P_{im}^{l} = \left(\frac{n}{n+1}\right)^{l}, \quad P_{podm}^{l}(s) = \left(\frac{n-1}{n}\right)^{s}.$$

В следующей теореме приводится конструкция совершенных имитостойких шифров на основе размещений.

Теорема 6.19. Пусть для шифра Σ_H выполнены следующие условия:

(i) $|U| = m$, $|V| = n$, $1 < m < n$, $r = A_n^m$ и все строки матрицы зашифрования опорного шифра Σ являются размещениями из n элементов множества V по m.

(ii) распределение вероятностей $P(\mathbb{N}_r)$ равномерно.

Тогда шифр Σ_H является совершенным, причем:

$$P_{im}^{l} = \left(\frac{m}{n}\right)^{l}, \quad P_{podm}^{l}(s) = \left(\frac{m-1}{n-1}\right)^{s}.$$

Доказательство. Пусть $u \in U$, $v \in V$. Так как $|\mathbb{N}_r(u,v)| = A_{n-1}^{m-1}$, то шифр Σ_H является совершенным.

Пусть $v, \widetilde{v} \in V$, $v \neq \widetilde{v}$. Тогда:

$$P(\mathbb{N}_r(v)) = \frac{|\mathbb{N}_r(v)|}{|\mathbb{N}_r|} = \frac{mA_{n-1}^{m-1}}{A_n^m} = \frac{m}{n}.$$

$$P(\mathbb{N}_r(\widetilde{v}) \mid \mathbb{N}_r(v)) = \frac{|\mathbb{N}_r(\widetilde{v}, v)|}{|\mathbb{N}_r(v)|} = \frac{A_m^2 A_{n-2}^{m-2}}{mA_{n-1}^{m-1}} = \frac{m-1}{n-1}.$$

Поэтому:

$$P_{im}^{l} = \left(\frac{m}{n}\right)^{l}, \quad P_{podm}^{l}(s) = \left(\frac{m-1}{n-1}\right)^{s}. \qquad \square$$

Пусть Σ_H — некоторый шифр замены с неограниченным ключом с опорным шифром $\Sigma = (U, \mathbb{N}_r, V, E, D)$, $|U| = n$, $|V| = s$, распределением вероятностей $P(\mathbb{N}_r)$ для случайного генератора ψ_c и матрицей зашифрования A размера $r \times n$ над множеством V для опорного шифра Σ. При этом строки матрицы A пронумерованы элементами множества $\mathbb{N}_r$, а столбцы — элементами множества U. Пусть также для некоторого $\widetilde{r} \geqslant r$ имеется

случайный генератор $\widetilde{\psi}_c$ с распределением вероятностей $P(\mathbb{N}_{\widetilde{r}})$ и условием, что найдется такое разбиение множества $\mathbb{N}_{\widetilde{r}}$ на r непустых непересекающиеся подмножеств:

$$\mathbb{N}_{\widetilde{r}} = K_1 \cup K_2 \cup \ldots \cup K_r,$$

для которого выполнены равенства:

$$P_{\mathbb{N}_{\widetilde{r}}}(K_i) = \sum_{k \in K_i} P_{\mathbb{N}_{\widetilde{r}}}(k) = P_{\mathbb{N}_r}(i), \quad i = 1, \ldots, r.$$

Построим шифр замены с неограниченным ключом $\widetilde{\Sigma}_H$ со случайным генератором $\widetilde{\psi}_c$ и опорным шифром $\widetilde{\Sigma} = (U, \mathbb{N}_{\widetilde{r}}, V, \widetilde{E}, \widetilde{D})$ со значениями U и V, как в опорном шифре Σ. Для этого необходимо определить множество правил зашифрования $\widetilde{E}$ и множество правил расшифрования $\widetilde{D}$. $\widetilde{E}$ и $\widetilde{D}$ определим с помощью матрицы зашифрования B размера $\widetilde{r} \times n$ над множеством V, в которой строки пронумерованы элементами множества $\mathbb{N}_{\widetilde{r}}$, а столбцы — элементами множества U, следующим образом: j-ю строку матрицы A продублируем $|K_j|$ раз, $j = 1, \ldots, r$, и из всех полученных (продублированных) строк составим матрицу B.

Предложение 6.6. Если один из шифров Σ_H или $\widetilde{\Sigma}_H$ является совершенным, то другой также будет являться совершенным. Более того, вероятности успехов имитации и успехов подмены данных шифров соответственно равны.

Доказательство следует из леммы 6.10 и определения понятий имитации и подмены. □

Заметим, что совершенные имитостойкие шифры можно строить не только для случая, когда $P(\mathbb{N}_r)$ равномерно. Пусть:

$$\mathbb{N}_r = K_1 \cup K_2 \cup \ldots \cup K_s \tag{6.21}$$

— разбиение множества $\mathbb{N}_r$ на непустые непересекающиеся подмножества с условием, что:

$$P_{\mathbb{N}_r}(K_i) = \sum_{k \in K_i} P_{\mathbb{N}_r}(k) = \frac{1}{s}, \quad i = 1, \ldots, s. \tag{6.22}$$

Пусть $U = \{u_1, \ldots, u_n\}$, A — матрица размера $s \times n$, $1 < n < s$, над множеством $V = \{v_1, \ldots, v_s\}$ вида:

v_1	v_2	$\ldots$	v_n
v_2	v_3	$\ldots$	v_{n+1}
$\ldots$	$\ldots$	$\ldots$	$\ldots$
v_s	v_1	$\ldots$	v_{n-1}

в которой каждый следующий столбец является циклическим сдвигом на одну позицию предыдущего столбца. Понятно, что данная матрица является латинским прямоугольником. Как и перед предложением 6.6, на основе матрицы A построим матрицу зашифрования B размера $r \times n$ над множеством V для опорного шифра Σ.

Предложение 6.7. Полученный шифр Σ_H будет являться совершенным, причем будут выполнены следующие равенства:

$$P^l_{im} = \left(\frac{n}{s}\right)^l, \quad P^l_{podm}(t) = \left(\frac{n-1}{n}\right)^t.$$

Доказательство следует из предложений 6.1, 6.2 и 6.6. □

Пример 6.11. Пусть:

$$U = \{u_1, u_2\}, \quad V = \{v_1, v_2, v_3\}, \quad \mathbb{N}_5 = \{1, 2, 3, 4, 5\}$$

и распределение вероятностей на множестве $\mathbb{N}_5$ имеет вид:

$\mathbb{N}_5$	1	2	3	4	5
$P(\mathbb{N}_5)$	1/15	4/15	1/12	1/4	1/3

В этом случае существует разбиение вида (6.21) с условием (6.22):

$$K_1 = \{1, 2\}, \ K_2 = \{3, 4\}, \ K_3 = \{5\},$$
$$P_{\mathbb{N}_5}(K_1) = P_{\mathbb{N}_5}(K_2) = P_{\mathbb{N}_5}(K_3) = \frac{1}{3}.$$

Сначала составим матрицу A :

$\mathbb{N}_3 \diagdown U$	u_1	u_2
1	v_1	v_2
2	v_2	v_3
3	v_3	v_1

которая является латинским прямоугольником, а на ее основе составим матрицу B :

$\mathbb{N}_5 \diagdown U$	u_1	u_2
1	v_1	v_2
2	v_1	v_2
3	v_2	v_3
4	v_2	v_3
5	v_3	v_1

По предложению 6.7 для данных U, V, $\mathbb{N}_5$, $P(\mathbb{N}_5)$ и матрицы зашифрования B для опорного шифра полученный шифр Σ_H будет являться совершенным, причем:

$$P_{im}^{l} = \left(\frac{2}{3}\right)^{l}, \quad P_{podm}^{l}(t) = \left(\frac{1}{2}\right)^{t}.$$

Пусть теперь:

$$\mathbb{N}_r = K_1 \cup K_2 \cup \ldots \cup K_{s^2-s} \tag{6.23}$$

— разбиение множества $\mathbb{N}_r$ с условием, что:

$$P_{\mathbb{N}_r}(K_i) = \sum_{k \in K_i} P_{\mathbb{N}_r}(k) = \frac{1}{s^2 - s}, \quad i = 1, \ldots, s^2 - s. \tag{6.24}$$

Обозначим через $A_j = A_j(s, 2)$ матрицу размера $s \times 2$ над множеством $V = \{v_1, \ldots, v_s\}$, имеющую такой вид:

$$A_j = \begin{pmatrix} v_1 & v_2 & \ldots & v_s \\ v_{T^j(1)} & v_{T^j(2)} & \ldots & v_{T^j(s)} \end{pmatrix}^{T}, \quad j = 1, \ldots, s-1.$$

Из матриц A_j, $j = 1, \ldots, s-1$, составим матрицу A размера $(s^2-s) \times 2$ путем последовательной графической записи матриц $A_1, \ldots, A_{s-1}$ одной под другой. Теперь на основе матрицы A построим матрицу зашифрования B размера $r \times 2$ для опорного шифра указанным выше способом.

Предложение 6.8. Полученный шифр Σ_H будет являться совершенным, причем будут выполнены следующие равенства:

$$P_{im}^{l} = \left(\frac{2}{s}\right)^{l}, \quad P_{podm}^{l}(t) = \left(\frac{1}{s-1}\right)^{t}.$$

Доказательство следует из предложений 6.5 и 6.6. □

6.11. Совершенные имитостойкие шифры на основе ортогональных таблиц

Напомним несколько важных определений.

Определение 6.11. *Латинским квадратом* s-го порядка над множеством $Y = \{y_1, \ldots, y_s\}$ называется таблица размера $s \times s$, заполненная элементами множества Y таким образом, что в каждой строке и в каждом столбце каждый элемент встречается ровно один раз.

Определение 6.12. Две матрицы $A = (a_{ij})$ и $B = (b_{ij})$ над множеством $Y = \{y_1, \ldots, y_s\}$ называются *ортогональными*, если все упорядоченные пары (a_{ij}, b_{ij}) различны.

Определение 6.13. *Ортогональной таблицей* $OA(s, n)$ над множеством $Y = \{y_1, \ldots, y_s\}$ называется матрица порядка $s^2 \times n$ над множеством Y с тем условием, что для любых двух столбцов данной матрицы каждая из пар $(y_i, y_j) \in Y \times Y$ встречается ровно один раз. Существование ортогональной таблицы $OA(s, n)$ над множеством Y эквивалентно существованию n попарно ортогональных квадратных матриц порядка s над множеством Y.

Нам понадобится следующая теорема.

Теорема 6.20 (**Bose**)**.** Для любого простого p и натурального d существуют $p^d - 1$ ортогональных латинских квадратов.

Доказательство. Пусть $F = GF(p^d)$ — конечное поле (теорема 2.16). Обозначим через A_k ($k \in F^*$) квадратную матрицу порядка p^d над полем F, в которой строки и столбцы пронумерованы элементами поля F, и каждый элемент матрицы A_k, находящийся на пересечении строки с номером x и столбца с номером y, вычисляется по правилу $kx + y$, $x, y \in F$. Нетрудно видеть, что для любого $k \in F^*$ матрица A_k является латинским квадратом.

Покажем, что для любых $k, \widetilde{k} \in F^*$, $k \neq \widetilde{k}$, матрицы A_k и $A_{\widetilde{k}}$ ортогональны. Предположим, что для некоторых $k, \widetilde{k}$ это не

верно. Тогда найдутся такие $x_1, x_2, y_1, y_2 \in F$, где либо $x_1 \neq x_2$, либо $y_1 \neq y_2$, для которых выполнены равенства:

$$kx_1 + y_1 = kx_2 + y_2, \quad \widetilde{k}x_1 + y_1 = \widetilde{k}x_2 + y_2.$$

Преобразуем данные равенства:

$$k(x_1 - x_2) = y_2 - y_1, \quad \widetilde{k}(x_1 - x_2) = y_2 - y_1. \tag{6.25}$$

1. Пусть $x_1 \neq x_2$. Из (6.25) следует, что:

$$k(x_1 - x_2) = \widetilde{k}(x_1 - x_2).$$

Поэтому $(k-\widetilde{k})(x_1-x_2) = 0$. Так как в поле нет делителей нуля, то $k = \widetilde{k}$, что противоречит предположению $k \neq \widetilde{k}$.

2. Пусть $y_1 \neq y_2$. Из (6.25) следует, что:

$$\frac{y_2 - y_1}{k} = \frac{y_2 - y_1}{\widetilde{k}}.$$

Поэтому $(k-\widetilde{k})(y_1-y_2) = 0$. Снова приходим к противоречию. □

Пример 6.12. Пусть в теореме 6.20 $p = 2$, $d = 2$. Построим три ортогональных латинских квадрата порядка 4. Пусть α — алгебраический элемент поля $GF(2^2)$ степени два над полем $\mathbb{Z}_2$ с неприводимым многочленом $x^2 - x - 1$ (см. пример 2.27). Все элементы поля $GF(4)$ можно представить в виде двоичных наборов длины два со следующим соответствием:

$$0 \to 00, \quad 1 \to 01, \quad \alpha \to 10, \quad \alpha + 1 \to 11.$$

Матрицы A_{01}, A_{10}, A_{11} из теоремы 6.20 будут соответственно иметь такой вид:

$$\begin{pmatrix} 00 & 01 & 10 & 11 \\ 01 & 00 & 11 & 10 \\ 10 & 11 & 00 & 01 \\ 11 & 10 & 01 & 00 \end{pmatrix}, \begin{pmatrix} 00 & 01 & 10 & 11 \\ 10 & 11 & 00 & 01 \\ 11 & 10 & 01 & 00 \\ 01 & 00 & 11 & 10 \end{pmatrix}, \begin{pmatrix} 00 & 01 & 10 & 11 \\ 11 & 10 & 01 & 00 \\ 01 & 00 & 11 & 10 \\ 10 & 11 & 00 & 01 \end{pmatrix}.$$

Ортогональная таблица легко получается из ортогональных латинских квадратов и наоборот.

Пример 6.13. Построим ортогональную таблицу $OA(4,3)$ над множеством $GF(4)$ на основе ортогональных латинских квадратов из примера 6.12. Для этого все строки матрицы A_{01} выпишем построчно. Под полученной строкой построчно выпишем все строки матрицы A_{10}. То же самое проделаем и с матрицей A_{11}. Транспонировав полученную матрицу, будем иметь ортогональную таблицу $OA(4,3)$ над полем $GF(4)$:

$$\begin{pmatrix} 00 & 01 & 10 & 11 & 01 & 00 & 11 & 10 & 10 & 11 & 00 & 01 & 11 & 10 & 01 & 00 \\ 00 & 01 & 10 & 11 & 10 & 11 & 00 & 01 & 11 & 10 & 01 & 00 & 01 & 00 & 11 & 10 \\ 00 & 01 & 10 & 11 & 11 & 10 & 01 & 00 & 01 & 00 & 11 & 10 & 10 & 11 & 00 & 01 \end{pmatrix}^T.$$

Пусть $OA(s,n)$ — ортогональная таблица над множеством $V = \{v_1, \ldots, v_s\}$, где s — степень простого числа, $1 < n < s$, в которой i-я строка содержит только элемент v_i, $i = 1, \ldots, s$ (теорема 6.20). Вычеркнем из таблицы $OA(s,n)$ первые s строк и обозначим полученную таблицу через $A(s,n)$. Понятно, что таблица $A(s,n)$ имеет размерность $(s^2 - s) \times n$, в каждой строке нет повторяющихся элементов, а каждый столбец содержит ровно $s-1$ экземпляров элемента v_i, $i = 1, \ldots, s$.

Теорема 6.21. Пусть для шифра Σ_H выполнены следующие условия:

(i) $|U| = n$, $|V| = s$, $1 < n < s$, $r = s^2 - s$;

(ii) матрица зашифрования опорного шифра представляет собой таблицу вида $A(s,n)$;

(iii) распределение вероятностей $P(\mathbb{N}_r)$ является равномерным.

Тогда шифр Σ_H является совершенным и для любого l выполнены следующие равенства:

$$P^l_{im} = \left(\frac{n}{s}\right)^l, \quad P^l_{podm}(t) = \left(\frac{n-1}{s-1}\right)^t,$$

т.е. $P^l_{im} \to 0$ при $l \to \infty$, $P^l_{podm}(t) \to 0$ при $t \to \infty$.

Доказательство. Совершенность шифра Σ_H следует из теоремы 6.13.

Пусть $v \in V$. Тогда:

$$P(\mathbb{N}_r(v)) = \frac{n(s-1)}{s(s-1)} = \frac{n}{s},$$

поэтому:

$$P_{im}^{l} = \left(\frac{n}{s}\right)^{l}.$$

Пусть $v, \widetilde{v} \in V$, $v \neq \widetilde{v}$. Тогда:

$$P(\mathbb{N}_r(\widetilde{v}) \mid \mathbb{N}_r(v)) = \frac{2C_n^2}{n(s-1)} = \frac{n-1}{s-1},$$

поэтому:

$$P_{podm}^{l}(t) = \left(\frac{n-1}{s-1}\right)^{t}. \qquad \square$$

Пример 6.14. На основе примера 6.13 построим совершенный имитостойкий шифр Σ_H со следующими характеристиками: $|U| = 3$, $|V| = 4$, $r = 12$ и следующей матрицей зашифрования:

$\mathbb{N}_r \backslash U$	u_1	u_2	u_3
1	01	10	11
2	00	11	10
3	11	00	01
4	10	01	00
5	10	11	01
6	11	10	00
7	00	01	11
8	01	00	10
9	11	01	10
10	10	00	11
11	01	11	00
12	00	10	01

Если распределение вероятностей $P(\mathbb{N}_r)$ является равномерным, то шифр Σ_H будет являться совершенным, причем:

$$P_{im}^{l} = \left(\frac{3}{4}\right)^{l}, \quad P_{podm}^{l}(t) = \left(\frac{2}{3}\right)^{t}.$$

Глава 7. Шифры, не распространяющие искажений

Передаваемое по каналу связи шифрованное сообщение может подвергнуться как целенаправленным искажениям злоумышленников, так и искажениям, причиной которых могут являться помехи в самом канале связи. Искажения могут привести к потере части или даже всего открытого текста, так как расшифрование искаженного шифрованного текста может привести к непредсказуемым результатам. Нас будут интересовать шифры, которые не распространяют искажения при расшифровании. Ограничимся рассмотрением эндоморфных шифров и таких искажений, которые заменяют символы алфавита символами того же алфавита либо приводят к потере или появлению дополнительных символов алфавита.

В данной главе под шифром Σ_A будем понимать определение 6.1, где множества X, Y и K не обязательно являются конечными. Основой результатов данной главы является работа [3].

7.1. Шифры, не распространяющие искажений типа замены знаков

Пусть A — некоторый конечный алфавит,

$$X = Y = \bigcup_{l=1}^{\infty} A^l.$$

Рассмотрим шифры, которые не изменяют длины сообщения при шифровании, т.е. такие шифры $\Sigma_A = (X, K, Y, E, D)$, что для любого $l \in \mathbb{N}$, любого $x \in A^l$ и любого $k \in K$ следует, что $E_k(x) \in A^l$. Поэтому $E_k(A^l) \subseteq A^l$ для любого $l \in \mathbb{N}$ и $k \in K$.

Так как все правила зашифрования E_k являются инъектив-

ными отображениями множества X в Y, то, с учетом того, что $X = Y$, все E_k будут являться также биективными преобразованиями множества X. В частности, $E_k(A^l) = A^l$ для любого $l \in \mathbb{N}$ и $k \in K$.

В A^l для любого $l \in \mathbb{N}$ введем метрику Хэмминга, определенную следующей формулой:

$$\rho(x, y) = \sum_{i=1}^{l} \delta(x_i, y_i),$$

где $x = x_1 \dots x_l$, $y = y_1 \dots y_l \in A^l$, причем:

$$\delta(x_i, y_i) = \begin{cases} 1, & x_i \neq y_i, \\ 0, & x_i = y_i. \end{cases}$$

Пусть $r > 0$ и $x \in A^l$. Определим шар радиуса r с центром в x как множество:

$$S_r(x) = \{y \in A^l \mid \rho(x, y) \leqslant r\}.$$

Определение 7.1. Говорят, что шифр Σ_A не распространяет искажений типа замены знаков, если для любого $l \in \mathbb{N}$, любых $x, y \in A^l$ и $k \in K$ выполнено неравенство:

$$\rho(D_k(x), D_k(y)) \leqslant \rho(x, y).$$

Данное неравенство означает следующее. Пусть x — некоторое шифрованное сообщение. Изменив в нем некоторые символы, получим искаженное сообщение y. Тогда при расшифровании искаженного шифрованного сообщения y получится сообщение $D_k(y)$, которое содержит не более $\rho(x, y)$ искажений относительно исходного сообщения $D_k(x)$, т.е. шифр Σ_A является помехоустойчивым.

Лемма 7.1. Для эндоморфного шифра Σ_A, не изменяющего длины сообщений, следующие условия эквивалентны.

(i) Шифр Σ_A не распространяет искажений типа замены знаков.

(ii) Для любого $l \in \mathbb{N}$, любых $x, y \in A^l$ и $k \in K$ выполнено равенство:

$$\rho(D_k(x), D_k(y)) = \rho(x, y). \tag{7.1}$$

(iii) Для любого $l \in \mathbb{N}$, любых $x, y \in A^l$ и $k \in K$ выполнено равенство:

$$\rho(E_k(x), E_k(y)) = \rho(x, y).$$

Доказательство. Очевидно, что если выполняется равенство (7.1), то шифр Σ_A не распространяет искажений типа замены знаков, поэтому из (ii) следует (i).

Докажем в обратную сторону. Зафиксируем некоторое $l \in \mathbb{N}$ и некоторый элемент $k \in K$. Рассмотрим следующее преобразование f множества $A^l \times A^l$:

$$f(x, y) = (D_k(x), D_k(y)), \quad (x, y) \in A^l \times A^l.$$

Так как D_k является биективным преобразованием множества A^l, то преобразование f множества $A^l \times A^l$ также является биективным. Поэтому если (x, y) пробегает все множество $A^l \times A^l$, то $(D_k(x), D_k(y))$ также пробегает все множество $A^l \times A^l$. Следовательно:

$$\sum_{(x,y)\in A^l\times A^l} \rho(D_k(x), D_k(y)) = \sum_{(x,y)\in A^l\times A^l} \rho(x, y).$$

Из данного равенства следует, что:

$$\sum_{(x,y)\in A^l\times A^l} [\rho(x, y) - \rho(D_k(x), D_k(y))] = 0.$$

Так как все слагаемые последней суммы неотрицательны и вся сумма равна 0, то это возможно лишь в том случае, когда выполнено условие (7.1). Следовательно, условие (i) влечет условие (ii).

Пусть выполнено условие (ii). Тогда:

$$\rho\big(E_k(x), E_k(y)\big) = \rho\big(D_k(E_k(x)), D_k(E_k(y))\big) = \rho(x, y).$$

Таким образом, из (ii) следует (iii).

Обратно, пусть выполнено условие (iii). Тогда:

$$\rho\big(D_k(x), D_k(y)\big) = \rho\big(E_k(D_k(x)), E_k(D_k(y))\big) = \rho(x, y).$$

Поэтому условие (iii) влечет условие (ii). □

Исходя из данной леммы, введем следующее определение.

Определение 7.2. Отображение $f : A^l \to A^l$ называется изометрией, если для любых $x, y \in A^l$ выполнено равенство:

$$\rho(f(x), f(y)) = \rho(x, y).$$

Заметим, что из определения 7.2 следует, что f является биективным преобразованием множества A^l.

Лемма 7.1 показывает, что если эндоморфный шифр Σ_A не распространяет искажений типа замены знаков, то множество правил зашифрования E состоит из изометрий.

Зафиксируем произвольным образом $\overline{a} = a_1 \dots a_l \in A^l$. Обозначим:

$$S_1^i(\overline{a}) = \{a_1 \dots a_{i-1} a a_{i+1} \dots a_l \mid a \in A\}.$$

Заметим, что:

$$\bigcup_{i=1}^{l} S_1^i(\overline{a}) = S_1(\overline{a}), \qquad \bigcap_{i=1}^{l} S_1^i(\overline{a}) = \begin{cases} A, & l = 1, \\ \{\overline{a}\}, & l > 1. \end{cases} \tag{7.2}$$

Лемма 7.2. Пусть $f : A^l \to A^l$ — изометрия и $\overline{a} = a_1 \dots a_l$ — некоторый элемент из A^l. Тогда будут выполнены следующие условия.

1. Для любого $r \geqslant 1$ справедливо равенство:

$$f(S_r(\overline{a})) = S_r(f(\overline{a})). \tag{7.3}$$

2. Для любого $i = 1, 2, \dots, l$ найдется такое j, $1 \leqslant j \leqslant l$, зависящее от значения i, что будет выполнено равенство:

$$f(S_1^i(\overline{a})) = S_1^j(f(\overline{a})). \tag{7.4}$$

Доказательство. 1. Пусть $r \geqslant 1$ — некоторое число. Покажем, что выполнено включение $f(S_r(\overline{a})) \subseteq S_r(f(\overline{a}))$. Пусть $\overline{y} \in f(S_r(\overline{a}))$. Тогда найдется такой элемент $\overline{x} \in S_r(\overline{a})$, что $f(\overline{x}) = \overline{y}$. При этом:

$$\rho(f(\overline{a}), \overline{y}) = \rho(f(\overline{a}), f(\overline{x})) = \rho(\overline{a}, \overline{x}) \leqslant r.$$

Следовательно, $\overline{y} \in S_r(f(\overline{a}))$, что показывает включение:

$$f(S_r(\overline{a})) \subseteq S_r(f(\overline{a})).$$

Поскольку:

$$|S_r(\overline{a})| = |f(S_r(\overline{a}))| = |S_r(f(\overline{a}))|,$$

то выполнено равенство (7.3).

2. Чтобы доказать равенство (7.4), достаточно показать, что выполнено вложение:

$$f(S_1^i(\overline{a})) \subseteq S_1^j(f(\overline{a})),$$

так как $|f(S_1^i(\overline{a}))| = |S_1^j(f(\overline{a}))|$ для любых $1 \leqslant i, j \leqslant l$.

Предположим противное. Пусть существует такое i_0, что для любого $j = 1, 2, \ldots, l$:

$$f(S_1^{i_0}(\overline{a})) \not\subseteq S_1^j(f(\overline{a})).$$

Заметим, что из (7.2) и пункта 1 данной леммы следует, что:

$$f\left(\bigcup_{i=1}^{l} S_1^i(\overline{a})\right) = \bigcup_{j=1}^{l} S_1^j(f(\overline{a})).$$

Поэтому найдутся такие $\overline{x}_1, \overline{x}_2 \in S_1^{i_0}(\overline{a})$, причем $\overline{x}_1 \neq \overline{a}$, $\overline{x}_2 \neq \overline{a}$, и такие $j_1 < j_2$, что:

$$\overline{y}_1 = f(\overline{x}_1) \in S_1^{j_1}(f(\overline{a})), \quad \overline{y}_2 = f(\overline{x}_2) \in S_1^{j_2}(f(\overline{a})).$$

Пусть $f(\overline{a}) = c_1 \ldots c_l$. Тогда:

$$\overline{y}_1 = c_1 \ldots c_{j_1-1} d c_{j_1+1} \ldots c_{j_2-1} c_{j_2} c_{j_2+1} \ldots c_l,$$

$$\overline{y}_2 = c_1 \ldots c_{j_1-1} c_{j_1} c_{j_1+1} \ldots c_{j_2-1} e c_{j_2+1} \ldots c_l,$$

причем $d \neq c_{j_1}$ и $e \neq c_{j_2}$. Поэтому $\rho(\overline{y}_1, \overline{y}_2) = 2$. Но:

$$\rho(\overline{y}_1, \overline{y}_2) = \rho(f(\overline{x}_1), f(\overline{x}_2)) = \rho(\overline{x}_1, \overline{x}_2) = 1.$$

Противоречие. □

Лемма 7.3. Пусть $f, g : A^l \to A^l$ — изометрии. Тогда:

(i) если для некоторого $\overline{a} \in A^l$ выполнено равенство $f|_{S_1(\overline{a})} = id|_{S_1(\overline{a})}$, где id — тождественное отображение, то $f = id$ на всем множестве A^l;

(ii) если для некоторого $\overline{a} \in A^l$ выполнено равенство $f|_{S_1(\overline{a})} = g|_{S_1(\overline{a})}$, то $f = g$ на всем множестве A^l.

Доказательство. (i) Рассмотрим последовательность вложенных шаров:

$$S_1(\overline{a}) \subseteq S_2(\overline{a}) \subseteq \ldots \subseteq S_n(\overline{a}) \subseteq \ldots$$

Заметим, что данная последовательность является стабилизирующейся, так как для любого $n \geqslant l$ выполняется равенство $S_n(\overline{a}) = A^l$. С помощью индукции покажем, что для любого $n \geqslant 1$ сужение отображения f на множество $S_n(\overline{a})$ равно тождественной функции на множестве $S_n(\overline{a})$:

$$f|_{S_n(\overline{a})} = id|_{S_n(\overline{a})},$$

Из условия (i) леммы база индукции при $n = 1$ выполняется.

Предположим, что для любого $k < n$ сужение отображения f на множество $S_k(\overline{a})$ является тождественным отображением. Покажем, что $f|_{S_n(\overline{a})} = id|_{S_n(\overline{a})}$, т.е. для любого $\overline{b} \in S_n(\overline{a})$ выполнено равенство $f(\overline{b}) = \overline{b}$. Не ограничивая общность, рассмотрим в качестве $\overline{b}$ такой элемент:

$$\overline{b} = b_1 b_2 \ldots b_n a_{n+1} \ldots a_l.$$

Если $\rho(\overline{a}, \overline{b}) < n$, тогда по предположению индукции выполнено равенство $f(\overline{b}) = \overline{b}$. Поэтому рассмотрим случай $\rho(\overline{a}, \overline{b}) = n$, который означает, что $a_1 \neq b_1$, $a_2 \neq b_2$,..., $a_n \neq b_n$. Введем в рассмотрение следующие элементы шара $S_{n-1}(\overline{a})$:

$$\overline{b}_1 = a_1 b_2 \ldots b_n a_{n+1} \ldots a_l,$$

$$\overline{b}_2 = b_1 a_2 \ldots b_n a_{n+1} \ldots a_l,$$

$$\ldots$$

$$\overline{b}_n = b_1 b_2 \ldots a_n a_{n+1} \ldots a_l.$$

По предположению индукции $f(\overline{b}_i) = \overline{b}_i$ для любого $i = 1, \ldots, n$. Поэтому:

$$\rho(f(\overline{b}), \overline{b}_i) = \rho(f(\overline{b}), f(\overline{b}_i)) = \rho(\overline{b}, \overline{b}_i) = 1.$$

Следовательно:

$$f(\overline{b}) \in \bigcap_{i=1}^{n} S_1(\overline{b}_i).$$

При этом:

$$\bigcap_{i=1}^{n} S_1(\overline{b}_i) = \begin{cases} \{\overline{a}, \overline{b}\}, & n = 2, \\ \{\overline{b}\}, & n > 2. \end{cases}$$

Так как f является биективным преобразованием множества A^l и $f(\overline{a}) = \overline{a}$, то $f(\overline{b}) = \overline{b}$.

(ii) Пусть для некоторого $\overline{a} \in A^l$ выполнено $f|_{S_1(\overline{a})} = g|_{S_1(\overline{a})}$.

Так как f и g биективные преобразования множества A^l, являющиеся изометриями, то преобразование $f \circ g^{-1}$ также будет являться биективным преобразованием множества A^l и изометрией, причем:

$$(f \circ g^{-1})|_{S_1(\overline{a})} = id|_{S_1(\overline{a})}.$$

Поэтому из пункта (i) будет следовать, что:

$$f \circ g^{-1} = id$$

на всем множестве A^l. Следовательно, $f = g$ на множестве A^l. $\square$

Пусть $l \in \mathbb{N}$. Определим на множестве A^l такие преобразования:

$$\Pi_{j_1 \ldots j_l}(a_1 \ldots a_l) = a_{j_1} \ldots a_{j_l},$$
$$R(a_1 \ldots a_l) = R_1(a_1) \ldots R_l(a_l),$$

где $R_1, \ldots, R_l$ — некоторые подстановки множества A, $\Pi_{j_1 \ldots j_l}$ — некоторая перестановка:

$$\Pi_{j_1 \ldots j_l} = \begin{pmatrix} 1 & 2 & \ldots & l \\ j_1 & j_2 & \ldots & j_l \end{pmatrix}.$$

Теорема 7.1. (А.А.Марков) Отображение $E_k \in E$ является изометрией тогда и только тогда, когда для любого $l \in \mathbb{N}$:

$$E_k|_{A^l} = R \circ \Pi_{j_1 \ldots j_l}$$

для подходящих R и $\Pi_{j_1 \ldots j_l}$, где $E_k|_{A^l}$ — сужение отображения E_k на множество A^l.

Доказательство. Так как преобразования R и $\Pi_{j_1 \ldots j_l}$ являются изометриями, а композиция изометрий также является изометрией, то достаточность условия теоремы очевидна.

Докажем в обратную сторону. Зафиксируем произвольное значение $l \in \mathbb{N}$ и некоторый элемент $\overline{a} = a_1 \dots a_l \in A^l$. Пусть $E_k(a_1 \dots a_l) = c_1 \dots c_l$. Из второго пункта леммы 7.2 следует, что для любого $i = 1, 2, \dots, l$ найдется такое $j = j(i)$, что:

$$E_k : \{a_1 \dots a_{i-1} a a_{i+1} \dots a_l \mid a \in A\} \to \\ \{c_1 \dots c_{j-1} a c_{j+1} \dots c_l \mid a \in A\}.$$

Поэтому:

$$E_k(a_1 \dots a_{i-1} a a_{i+1} \dots a_l) = c_1 \dots c_{j-1} R_{j_i}(a) c_{j+1} \dots c_l, \qquad (7.5)$$

где R_{j_i} — некоторая подстановка множества A. Следовательно, сужение отображения E_k на множество $S_1(\overline{a})$ представимо в виде композиции преобразований R и $\Pi_{j_1 \dots j_l}$:

$$E_k|_{S_1(\overline{a})} = (R \circ \Pi_{j_1 \dots j_l})|_{S_1(\overline{a})},$$

где $R_1, \dots, R_l$ — преобразования, полученные в формуле (7.5). Так как композиция изометрий R и $\Pi_{j_1 \dots j_l}$ также является изометрией, то из последнего равенства в силу леммы 7.3 следует такое равенство:

$$E_k|_{A^l} = R \circ \Pi_{j_1 \dots j_l}. \qquad \square$$

Из теоремы А.А. Маркова следует, что в классе эндоморфных шифров, не изменяющих длины сообщений, не распространяют искажений типа замены знаков, например, шифры перестановки, поточные шифры однозначной замены (шифры замены с $U = V = A$, где U — множество шифрвеличин, V — множество шифробозначений), а также композиции шифров перестановки и замены.

7.2. Шифры, не распространяющие искажений типа пропуска знаков

Пусть A — некоторый конечный алфавит, $X = Y = \bigcup_{l=1}^{\infty} A^l$. В данном пункте, как и в предыдущем, будем рассматривать эндоморфные шифры $\Sigma_A = (X, K, Y, E, D)$.

Введем на множестве $X = Y$ бинарное отношение ε следующим образом. Пусть $x, y \in X$. $x \varepsilon y \Leftrightarrow$ слово y получено из x

путем удаления одного вхождения некоторой его буквы. Определим множество $\varepsilon(x)$ для некоторого $x \in X$ следующим образом:

$$\varepsilon(x) = \{y \in X \mid x\varepsilon y\}.$$

Например, если $x = x_1x_2x_3$, то $\varepsilon(x) = \{x_1x_2, x_1x_3, x_2x_3\}$.

Через ε^n будем понимать степень отношения ε : $x\varepsilon^n y$, где значение n меньше длины слова x, тогда и только тогда, когда слово y получено из x путем удаления n вхождений некоторых его букв. Можно также дать и эквивалентное определение для ε^n : $x\varepsilon^n y$, где значение n меньше длины слова x, тогда и только тогда, когда найдутся такие z_1, z_2,. . . , $z_{n-1} \in X$, что:

$$x \ \varepsilon \ z_1 \ \varepsilon \ z_2 \ \varepsilon \ \ldots \ \varepsilon \ z_{n-1} \ \varepsilon \ y.$$

Заметим, что $x\varepsilon^0 y \Leftrightarrow x = y$.

Определение 7.3. Будем говорить, что шифр Σ_A не распространяет искажений типа пропуска знаков, если для любых $x, y \in Y$, любого $k \in K$ и любого натурального n, меньшего длины слова x, найдется такое число m, $0 \leqslant m \leqslant n$, что из условия $x\varepsilon^n y$ следует $D_k(x)\varepsilon^m D_k(y)$.

Лемма 7.4. Так как в определении 7.3 число $n > 0$, то из этого всегда будет следовать, что число m также больше нуля.

Доказательство. Пусть для некоторых $x, y \in Y$ из условия $x\varepsilon^n y$ следует $D_k(x)\varepsilon^0 D_k(y)$, где n — некоторое положительное число, меньшее длины слова x, и k — некоторый элемент множества K. Так как $x\varepsilon^n y$, то $x \neq y$, а из условия $D_k(x)\varepsilon^0 D_k(y)$ следует, что $D_k(x) = D_k(y)$, что противоречит инъективности отображения D_k. □

Лемма 7.5. Если эндоморфный шифр Σ_A не распространяет искажений типа пропуска знаков, то он не изменяет длины сообщений при шифровании.

Доказательство. Пусть выполнено условие леммы для некоторого шифра Σ_A. Покажем, что для любого $k \in K$ и любого $l \in \mathbb{N}$ выполнено включение $E_k(A^l) \subseteq A^l$.

Фиксируем произвольное $k \in K$. Рассмотрим сначала случай $l = 1$. Предположим, что $E_k(A) \not\subseteq A$, т.е. найдется такое $a \in A$, что $E_k(a) = b_1 \dots b_t \in A^t$, где $t > 1$. Так как $b_1 \dots b_t \varepsilon b_1 \dots b_{t-1}$, то $D_k(b_1 \dots b_t) \varepsilon D_k(b_1 \dots b_{t-1})$. Но $D_k(b_1 \dots b_t) = a \in A$, а длина слова $D_k(b_1 \dots b_{t-1})$ не меньше единицы, поэтому, с учетом леммы 7.4, пришли к противоречию. Следовательно, $E_k(A) \subseteq A$.

Предположим, что для всех $t < l$, где $l > 1$, выполнено включение $E_k(A^t) \subseteq A^t$ для любого $k \in K$. Покажем, что тогда $E_k(A^l) \subseteq A^l$. Предположим, что это не так для некоторого $k \in K$. Пусть $E_k(A^l) \not\subseteq A^l$, т.е. найдется такое $x \in A^l$, что $E_k(x) = b_1 \dots b_s$, причем $s \neq l$. Заметим, что число s не может быть меньше чем l, так как $D_k(A^t) \subseteq A^t$ для всех $t < l$. Поэтому $s > l$. Из отношения $b_1 \dots b_s \varepsilon b_1 \dots b_{s-1}$ должно следовать отношение $D_k(b_1 \dots b_s) \varepsilon D_k(b_1 \dots b_{s-1})$, но длина слова $D_k(b_1 \dots b_s)$ равна l, а длина слова $D_k(b_1 \dots b_{s-1})$ не меньше l, что следует из предположения индукции, так как $s - 1 \geqslant l$. Поэтому пришли к противоречию. Следовательно $E_k(A^l) \subseteq A^l$. □

Лемма 7.6. Для эндоморфного шифра Σ_A следующие условия эквивалентны.

(i) Шифр Σ_A не распространяет искажений типа пропуска знаков.

(ii) Для любых $x, y \in Y$, любого $k \in K$ и любого натурального n, меньшего длины слова x, из условия $x \varepsilon^n y$ следует $D_k(x) \varepsilon^n D_k(y)$.

(iii) Для любых $x, y \in X$, любого $k \in K$ и любого натурального n, меньшего длины слова x, из условия $x \varepsilon^n y$ следует $E_k(x) \varepsilon^n E_k(y)$.

Доказательство. Очевидно, что из условия (ii) следует (i).

Докажем в обратную сторону. Предположим, что выполнено условие (i). Пусть $x, y \in Y$ и $x \varepsilon^n y$, т.е. найдутся такие z_1, z_2,. . . , $z_{n-1} \in Y$, что:

$$x \ \varepsilon \ z_1 \ \varepsilon \ z_2 \ \varepsilon \ \dots \ \varepsilon \ z_{n-1} \ \varepsilon \ y.$$

Тогда, учитывая лемму 7.4, имеем:

$$D_k(x) \ \varepsilon \ D_k(z_1) \ \varepsilon \ D_k(z_2) \ \varepsilon \ \dots \ \varepsilon \ D_k(z_{n-1}) \ \varepsilon \ D_k(y).$$

Следовательно $D_k(x)\varepsilon^n D_k(y)$. Поэтому условие (i) влечет условие (ii).

Пусть выполнено условие (ii). Зафиксируем некоторое значение $k \in K$. Пусть $x, y \in X$. Так как шифр Σ_A сохраняет длины сообщений при шифровании (лемма 7.5), то найдется такое число m, что:

$$D_k^m(x) = \underbrace{(D_k \circ \ldots \circ D_k)}_{m}(x) = x,$$

$$D_k^m(y) = \underbrace{(D_k \circ \ldots \circ D_k)}_{m}(y) = y.$$

Поэтому $E_k(x) = D_k^{m-1}(x)$ и $E_k(y) = D_k^{m-1}(y)$. Так как из отношения $x\varepsilon^n y$ следует $D_k(x)\varepsilon^n D_k(y)$, из которого, в свою очередь, следует $D_k^2(x)\varepsilon^n D_k^2(y)$ и т.д., то:

$$E_k(x) = D_k^{m-1}(x)\ \varepsilon^n\ D_k^{m-1}(y) = E_k(y).$$

Таким образом, из условия (ii) следует (iii).

Аналогичным образом доказывается, что условие (iii) влечет условие (ii). □

Лемма 7.7. Пусть эндоморфный шифр Σ_A не распространяет искажений типа пропуска знаков. Тогда для любого $x \in X$ и любого $k \in K$ следует равенство:

$$E_k(\varepsilon(x)) = \varepsilon(E_k(x)).$$

Доказательство. Зафиксируем $x \in X$ и $k \in K$. Покажем сначала, что:

$$E_k(\varepsilon(x)) \subseteq \varepsilon(E_k(x)). \tag{7.6}$$

Пусть $y \in E_k(\varepsilon(x))$. Тогда найдется такой элемент $\widetilde{x} \in \varepsilon(x)$, что $y = E_k(\widetilde{x})$. Так как $x\varepsilon\widetilde{x}$, то $E_k(x)\varepsilon E_k(\widetilde{x}) = y$ (лемма 7.6). Следовательно, $y \in \varepsilon(E_k(x))$, что доказывает включение (7.6).

Аналогичным же образом получается, что:

$$D_k(\varepsilon(x)) \subseteq \varepsilon(D_k(x)),$$

из которого следует, что:

$$D_k\big(\varepsilon(E_k(x))\big) \subseteq \varepsilon\big(D_k(E_k(x))\big) = \varepsilon(x). \tag{7.7}$$

Из включения (7.7) следует такое включение:

$$\varepsilon(E_k(x)) = E_k\big(D_k(\varepsilon(E_k(x)))\big) \subseteq E_k(\varepsilon(x)). \qquad \square$$

Лемма 7.8. Пусть $\overline{x}, \overline{y} \in A^l$ для некоторого $l \geqslant 3$. Тогда из равенства $\varepsilon(\overline{x}) = \varepsilon(\overline{y})$ будет следовать равенство $\overline{x} = \overline{y}$.

Доказательство. Докажем сначала, что любое слово $\overline{x} \in A^l$, где $l \geqslant 3$, однозначно определяется множеством $\varepsilon(\overline{x})$.

Пусть a — первый символ слова $\overline{x}$. Тогда возможны 3 случая:

1) $\overline{x} = a^l = \underbrace{a \ldots a}_{l}$,
2) $\overline{x} = a^n b\overline{z}, \quad n \geqslant 2,\ b \in A,\ a \neq b$,
3) $\overline{x} = ab\overline{z}, \quad b \in A,\ a \neq b$.

В первом случае множество $\varepsilon(\overline{x})$ состоит из одного слова a^{l-1}.

Во втором случае все слова из $\varepsilon(\overline{x})$ начинаются с буквы a, причем одно из них есть $a^{n-1}b\overline{z}$, а все другие имеют начало a^n.

В третьем случае $\varepsilon(\overline{x})$ содержит слова $a\overline{z}$ и $b\overline{z}$, а все остальные слова (если они есть, т.е. $\overline{z} \neq b^{l-2}$) имеют начало ab.

Поэтому алгоритм восстановления слова $\overline{x}$ по множеству $\varepsilon(\overline{x})$ будет следующим.

1. Если $\varepsilon(\overline{x}) = \{a^{l-1}\}$, то $\overline{x} = a^l$.

2. Если все слова из $\varepsilon(\overline{x})$ начинаются на одну и ту же букву, например, a, то одним из них является $a^{n-1}b\overline{z}$, а все остальные имеют вид $a^n\overline{z}_i$. Тогда $\overline{x} = a^n b\overline{z}$.

3. Пусть все слова, кроме одного, множества $\varepsilon(\overline{x})$ начинаются на одну и ту же букву.

a) Если $\varepsilon(\overline{x}) = \{a\overline{z}, b\overline{z}\}$, то либо $\overline{z} = a^{l-2}$, либо $\overline{z} = b^{l-2}$. Поэтому если, например, $\overline{z} = b^{l-2}$, то $\overline{x} = ab^{l-1}$.

b) Если же $\varepsilon(\overline{x}) = \{a\overline{z}, b\overline{z}, ab\overline{z}_1, \ldots, ab\overline{z}_m\}$, то из множества $\varepsilon(\overline{x})$ возьмем единственное слово, не начинающееся на букву a, т.е. $b\overline{z}$, и добавим к нему в начало букву a, получаем $\overline{x} = ab\overline{z}$.

Таким образом, если $\varepsilon(\overline{x}) = \varepsilon(\overline{y})$ для некоторых $\overline{x}, \overline{y} \in A^l$, где $l \geqslant 3$, то $\overline{x} = \overline{y}$. $\square$

Обозначим через σ такое биективное преобразование множества X, для которого выполнено свойство:

$$\sigma(x_1 \ldots x_l) = \sigma(x_1) \ldots \sigma(x_l)$$

для любого $x = x_1 \dots x_l \in X$. Таким образом, преобразование σ достаточно задать на множестве A. Обозначим также через μ оператор обращения слов, т.е. если $x = x_1 \dots x_l$, то $\mu(x) = x_l \dots x_1$.

Теорема 7.2. Пусть $X = Y = \bigcup_{l=1}^{L} A^l$. Эндоморфный шифр Σ_A не распространяет искажений типа пропуска знаков тогда и только тогда, когда для любого $k \in K$ выполнены следующие условия:

1. Если $L = 2$, то для любого $x \in X$ либо:

$$E_k(x) = \sigma(x),$$

либо:

$$E_k(x) = \sigma(\mu(x)),$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A;

2. Если $L > 2$, то либо:

$$E_k = \sigma \text{ на всем множестве } X,$$

либо:

$$E_k = \sigma \circ \mu \text{ на всем множестве } X,$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A.

Доказательство. Очевидно, что отображения σ и μ не распространяют искажений типа пропуска знаков, поэтому достаточность условия теоремы очевидна.

Докажем в обратную сторону. Зафиксируем $k \in K$.

1. Пусть $L = 2$. Так как шифр Σ_A не изменяет длины слов при шифровании (лемма 7.5), то $E_k(A) = A$. Обозначим через σ — сужение отображения E_k на множество A. Покажем, что для любого $a_1a_2 \in A^2$ либо $E_k(a_1a_2) = \sigma(a_1)\sigma(a_2)$, либо $E_k(a_1a_2) = \sigma(a_2)\sigma(a_1)$.

Пусть $a_1 \neq a_2$ и $E_k(a_1a_2) = b_1b_2 \in A^2$. Из леммы 7.7 следует, что:

$$E_k(\varepsilon(a_1a_2)) = \varepsilon(E_k(a_1a_2)).$$

Так как:

$$E_k(\varepsilon(a_1a_2)) = E_k(\{a_1, a_2\}) = \{E_k(a_1), E_k(a_2)\} = \{\sigma(a_1), \sigma(a_2)\},$$

$$\varepsilon(E_k(a_1a_2)) = \varepsilon(b_1b_2) = \{b_1, b_2\},$$

то либо $\sigma(a_1) = b_1$ и $\sigma(a_2) = b_2$, либо $\sigma(a_1) = b_2$ и $\sigma(a_2) = b_1$.

Пусть $a_1 = a_2 = a$ и $E_k(aa) = b_1b_2$. Тогда:

$$E_k(\varepsilon(aa)) = E_k(\{a\}) = \{\sigma(a)\},$$

$$\varepsilon(E_k(aa)) = \varepsilon(b_1b_2) = \{b_1, b_2\}.$$

Следовательно, $b_1 = b_2 = \sigma(a)$.

2. Пусть $L > 2$. Из пункта 1 следует, что для любого $a_1a_2 \in A^2$ либо $E_k(a_1a_2) = \sigma(a_1)\sigma(a_2)$, либо $E_k(a_1a_2) = \sigma(a_2)\sigma(a_1)$, где $\sigma = E_k|_A$ — сужение отображения E_k на множество A. Покажем, что в случае $L > 2$ отображение E_k на всем множестве A^2 представимо либо в виде:

$$E_k|_{A^2} = \sigma,$$

либо в виде:

$$E_k|_{A^2} = \sigma \circ \mu.$$

Рассмотрим такой случай: пусть найдутся такие $a_1, a_2 \in A$, $a_1 \neq a_2$, что:

$$E_k(a_1a_2) = \sigma(a_1)\sigma(a_2).$$

Покажем, что тогда $E_k|_{A^2} = \sigma$. Для начала докажем, что в этом случае $E_k(a_1a) = \sigma(a_1)\sigma(a)$ для любого $a \in A$. Предположим, что это не так. Тогда найдется такое $a \in A$, $a \neq a_1$, $a \neq a_2$, что $E_k(a_1a) = \sigma(a)\sigma(a_1)$. Рассмотрим слово $a_1a_2a \in A^3$. Так как:

$$\varepsilon(a_1a_2a) = \{a_1a_2, a_1a, a_2a\},$$

то:

$$E_k(\varepsilon(a_1a_2a)) = \{\sigma(a_1)\sigma(a_2), \sigma(a)\sigma(a_1), E_k(a_2a)\}.$$

Поскольку $E_k(\varepsilon(a_1a_2a)) = \varepsilon(E_k(a_1a_2a))$ (лемма 7.7) и шифрованное слово $E_k(a_1a_2a)$ однозначно определяется по множеству $\varepsilon(E_k(a_1a_2a))$ (лемма 7.8), то либо слово $E_k(a_2a)$ должно начинаться на букву $\sigma(a_1)$, либо на букву $\sigma(a)$. Так как:

$$E_k(a_2a) \in \{\sigma(a_2)\sigma(a), \sigma(a)\sigma(a_2)\},$$

то $E_k(a_2a) = \sigma(a)\sigma(a_2)$ и поэтому $E_k(a_1a_2a) = \sigma(a)\sigma(a_1)\sigma(a_2)$. Теперь рассмотрим слово $a_1aa_2 \in A^3$. Так как:

$$\varepsilon(a_1aa_2) = \{a_1a, a_1a_2, aa_2\}$$

и $E_k(aa_2) = \sigma(a_2)\sigma(a)$, то:

$$E_k(\varepsilon(a_1aa_2)) = \{\sigma(a)\sigma(a_1), \sigma(a_1)\sigma(a_2), \sigma(a_2)\sigma(a)\}.$$

Но из полученного множества нельзя собрать слово, так как в данном множестве три слова начинаются на три различные буквы, а этого не может быть (лемма 7.8). Так как должно быть $E_k(\varepsilon(a_1aa_2)) = \varepsilon(E_k(a_1aa_2))$, то слова $E_k(a_1aa_2)$ не существует. Противоречие.

Таким же образом показывается, что $E_k(aa_2) = \sigma(a)\sigma(a_2)$ для любого $a \in A$. Пусть $ab \in A^2$. Тогда $E_k(a_1b) = \sigma(a_1)\sigma(b)$, а из данного равенства будет следовать равенство $E_k(ab) = \sigma(a)\sigma(b)$.

Совершенно аналогично доказывается, что если найдутся такие $a_1, a_2 \in A$, $a_1 \neq a_2$, что:

$$E_k(a_1a_2) = \sigma(a_2)\sigma(a_1),$$

то $E_k|_{A^2} = \sigma \circ \mu$.

Таким образом, условие пункта 2 доказано для всех слов из множества $A \cup A^2$. Предположим, что условие пункта 2 верно для всех слов из множества $A \cup \ldots \cup A^t$, где $t < l$. Докажем утверждение для случая $t = l$. Пусть для определенности $E_k|_{A\cup\ldots\cup A^t} = \sigma$. Зафиксируем $x \in A^l$. Из леммы 7.7 следует, что:

$$E_k(\varepsilon(x)) = \varepsilon(E_k(x)).$$

Исходя из предположения индукции и леммы 7.7, имеем:

$$E_k(\varepsilon(x)) = \sigma(\varepsilon(x)) = \varepsilon(\sigma(x)).$$

Из полученного равенства $\varepsilon(E_k(x)) = \varepsilon(\sigma(x))$ немедленно вытекает равенство $E_k(x) = \sigma(x)$ (лемма 7.8).

Если же $E_k|_{A^t} = \sigma \circ \mu$, то:

$$\varepsilon(E_k(x)) = E_k(\varepsilon(x)) = (\sigma \circ \mu)(\varepsilon(x)) = \varepsilon((\sigma \circ \mu)(x)).$$

Поэтому $E_k(x) = (\sigma \circ \mu)(x)$. В силу произвольности элемента $x \in A^l$ получаем справедливость утверждения пункта 2. $\square$

Следствие 7.1. Пусть $X = Y = \bigcup_{l=1}^{\infty} A^l$. Эндоморфный шифр Σ_A не распространяет искажений типа пропуска знаков тогда и только тогда, когда для любого $k \in K$ либо:

$$E_k = \sigma \text{ на всем множестве } X,$$

либо:

$$E_k = \sigma \circ \mu \text{ на всем множестве } X,$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A.

7.3. Шифры, не распространяющие искажений типа вставки знаков

Пусть A — некоторый конечный алфавит, $X = Y = \bigcup_{l=1}^{\infty} A^l$. В данном пункте рассматриваются эндоморфные шифры $\Sigma_A = (X, K, Y, E, D)$.

Введем на множестве X бинарное отношение ε следующим образом. Пусть $x, y \in Y$. $x\varepsilon y \Leftrightarrow$ слово x получено из y путем добавления одной буквы.

Как и прежде, определим множество $\varepsilon(x)$ для некоторого элемента $x \in X$ следующим образом:

$$\varepsilon(x) = \{y \in X \mid x\varepsilon y\}.$$

Определение 7.4. Будем говорить, что шифр Σ_A не распространяет искажений типа вставки знаков, если для любых $x, y \in Y$, любого $k \in K$ и любого натурального n найдется такое число $0 \leqslant m \leqslant n$, что из условия $x\varepsilon^n y$ следует $D_k(x)\varepsilon^m D_k(y)$.

Очевидно, что все леммы из предыдущего пункта будут справедливы и для шифров, не распространяющих искажений типа вставки знаков. Поэтому доказательство следующей теоремы аналогично доказательству теоремы 7.2.

Теорема 7.3. Пусть $X = Y = \bigcup_{i=1}^{L} A^i$. Эндоморфный шифр Σ_A не распространяет искажений типа вставки знаков тогда и только тогда, когда для любого $k \in K$ выполнены следующие условия:

1. Если $L = 2$, то для любого $x \in X$ либо:

$$E_k(x) = \sigma(x),$$

либо:

$$E_k(x) = \sigma(\mu(x)),$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A;

2. Если $L > 2$, то либо:

$$E_k = \sigma \text{ на всем множестве } X,$$

либо:

$$E_k = \sigma \circ \mu \text{ на всем множестве } X,$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A.

Следствие 7.2. Пусть $X = Y = \bigcup_{l=1}^{\infty} A^l$. Эндоморфный шифр Σ_A не распространяет искажений типа вставки знаков тогда и только тогда, когда для любого $k \in K$ либо:

$$E_k = \sigma \text{ на всем множестве } X,$$

либо:

$$E_k = \sigma \circ \mu \text{ на всем множестве } X,$$

где $\sigma = E_k|_A$ — сужение отображения E_k на множество A.

Из теорем 7.1, 7.2, 7.3 получается такое следствие.

Следствие 7.3. Пусть:

$$X = Y = \bigcup_{l=1}^{L} A^l, \ L > 2, \text{ либо } X = Y = \bigcup_{l=1}^{\infty} A^l.$$

Тогда если шифр Σ_A не распространяет искажений типа пропуска (вставки) знаков, то он также не распространяет искажений типа замены знаков.

Глава 8. Симметричные блочные шифры

8.1. Общие сведения

В блочных алгоритмах шифрования входная последовательность битов разбивается на участки определенной длины (например, 64 бита для удобства реализации на процессорах с внутренними регистрами длиною 32 или 64 бита), и преобразование каждого блока совершается отдельно. При этом выходная последовательность алгоритма симметричного блочного шифрования представляет собой блоки, длина которых совпадает с длиной входных блоков.

Если длина открытого текста не кратна длине входных блоков (n бит), то последний блок можно дополнить двоичным вектором $10\ldots0$ до необходимой длины. Если же длина последнего блока равна n, то к сообщению присоединяется дополнительный n-битовый блок $10\ldots0$. Такой способ позволяет однозначно распознать присоединенный набор бит $10\ldots0$ и отбросить его при расшифровании сообщения.

Можно также поступить и следующим образом. Неполный последний m-битовый $(0 \leqslant m < n)$ блок дополняем $n-m-8$ специальными битами, например, нулевыми. А на место оставшихся 8 бит записываем число $n-m$, равное дефициту битов последнего блока. Если при этом оказывается, что $n-m<8$, то к сообщению придется добавить дополнительный блок. Ситуация $m=0$ соответствует случаю, когда в сообщении нет неполного блока. В этом случае для однозначности распознавания неполных блоков к сообщению следует присоединить дополнительный блок, состоящий из $n-8$ специальных битов, а на место оставшихся 8 бит записать число n.

Широкое практическое применение блочных шифров обусловлено следующими факторами:

- возможностью эффективной программной реализации на современных аппаратно-программных средствах;
- высокой скоростью зашифрования и расшифрования как при аппаратной, так и при программной реализации (при этом скорость реализации блочных алгоритмов значительно превышает скорости реализации шифров с открытыми ключами и теоретически стойких шифров);
- высокой вычислительной стойкостью данных шифров (т.е. для вскрытия шифра понадобится, например, 10^{20} лет при использовании суперкомпьютера).

При этом шифр называется *вычислительно стойким* (при атаке по выбранному тексту), если для него не существуют алгоритма взлома, существенно более быстрого, чем прямой перебор ключей.

Для разработки вычислительно стойких алгоритмов используются два общих принципа: рассеивание и перемешивание.

Рассеиванием называется распространение влияния одного символа открытого текста на много символов шифртекста, которое приводит к сокрытию статистических свойств исходного сообщения. *Перемешивание* — преобразование исходного сообщения, в результате которого вероятные последовательности (например, биграммы, триграммы и т.д.) рассеиваются по всему сообщению, нарушая статистические свойства исходного открытого текста.

8.2. Итеративные блочные шифры

Рассмотрим конструкции блочных шифров из [43]. Пусть, как и прежде, K — конечное множество ключей. Обозначим через $Q(k)$ множество производных ключей, зависящих от ключа $k \in K$ (например, если k — 256-битный ключ, то в качестве $Q(k)$ можно рассматривать 8 32-битных подключей ключа k).

Такой же смысл придадим множеству $Q'(k)$, которое, возможно, имеет иную природу, нежели $Q(k)$. Для краткости записи будем писать Q и Q' вместо $Q(k)$ и $Q'(k)$. Пусть V_n — множество двоичных векторов длины n.

Пусть $\varphi(x, q)$ — биективное по x отображение $V_n \times Q \to V_n$, где $x \in V_n$, $q \in Q$, т.е. для любого фиксированного $q \in Q$ отображение $\varphi_q : V_n \to V_n$, определенное равенством:

$$\varphi_q(x) = \varphi(x, q),\ x \in V_n,$$

является биективным преобразованием множества V_n.

Рассмотрим также биективные по переменной x отображения $\alpha(x, q')$ и $\beta(x, q')$:

$$\alpha,\ \beta : V_n \times Q' \to V_n,$$

где $x \in V_n$, $q' \in Q'$.

Для реализации правила зашифрования E_k используются производные ключи $q_1, \ldots, q_r \in Q$, $q'_0, q'_{r+1} \in Q'$:

$$q_i = \theta_i(k), \quad i = 1, \ldots, r,$$

$$q'_0 = \theta'_0(k), \quad q'_{r+1} = \theta'_{r+1}(k),$$

где $\{\theta'_0, \theta_1, \ldots, \theta_r, \theta'_{r+1}\}$ — семейство функций, отображающее множество K во множество $Q' \times Q^r \times Q'$. Данное семейство называется *расписанием использования ключа шифрования*, число r называется *числом раундов шифрования.*

Правило зашифрования E_k можно представить в виде следующей композиции (преобразования применяются справа налево):

$$E_k(x) = (\beta_{q'_{r+1}} \circ \varphi_{q_r} \circ \ldots \circ \varphi_{q_1} \circ \alpha_{q'_0})(x), \quad x \in V_n. \tag{8.1}$$

Правило расшифрования D_k будет выглядеть следующим образом:

$$D_k(y) = (\alpha^{-1}_{q'_0} \circ \varphi^{-1}_{q_1} \circ \ldots \circ \varphi^{-1}_{q_r} \circ \beta^{-1}_{q'_{r+1}})(y), \quad y \in V_n. \tag{8.2}$$

Функция $\varphi(x, q)$ называется *раундовой функцией*, отображение $\alpha(x, q')$ — входным отображением, $\beta(x, q')$ — выходным отображением.

Величина q_i называется *i-ым раундовым ключом* алгоритма шифрования, q'_0 и q'_{r+1} — соответственно *ключами входного и выходного отображения.*

Часть алгоритма зашифрования, ограниченная применением подстановки φ_{q_i}, называется *i-ым раундом шифрования.* Если $\varphi_{q_i}(x') = y'$, то x' называется входным блоком, y' — выходным блоком i-го раунда шифрования.

Для удобства аппаратной реализации итеративный блочный шифр можно построить таким образом, чтобы обеспечивалась его обратимость.

Теорема 8.1. Если для любого $q' \in Q'$ подстановка $\beta_{q'}$ является обратной к $\alpha_{q'}$ (т.е. $\alpha_{q'}^{-1} = \beta_{q'}$) и для любого $q \in Q$ подстановка φ_q является инволюцией (т.е. $\varphi_q^{-1} = \varphi_q$), то итеративный блочный шифр обратим и правило расшифрования отличается от правила зашифрования только тем, что раундовые ключи используются в обратном порядке.

Доказательство. Пусть выполнены условия теоремы. Тогда выполнены следующие равенства:

$$\alpha_{q'_0}^{-1} = \beta_{q'_0}, \quad \beta_{q'_{r+1}}^{-1} = \alpha_{q'_{r+1}}, \quad \varphi_q^{-1} = \varphi_q$$

для любых $q'_0, q'_{r+1} \in Q'$, $q \in Q$. Поэтому равенство (8.2) будет иметь такой вид:

$$D_k(y) = (\beta_{q'_0} \circ \varphi_{q_1} \circ \ldots \circ \varphi_{q_r} \circ \alpha_{q'_{r+1}})(y), \quad y \in V_n.$$

Осталось сравнить полученное равенство с равенством (8.1). □

8.3. Шифры Фейстеля

Одним из первых способов построения раундовой функции были шифры Фейстеля. Шифр Фейстеля — итеративный блочный шифр, раундовая функция которого оперирует с левой и правой половинами входного блока x ($x = (x_1, x_2)$) и имеет вид:

$$\varphi(x, q) = \varphi((x_1, x_2), q) = (x_2, f(x_2, q) \oplus x_1), \qquad (8.3)$$

где $f : V_m \times Q \to V_m$ $(n = 2m)$. Запишем для удобства последнее равенство в такой форме:

$$\left[\begin{array}{l} y_1 = x_2, \\ y_2 = f(x_2, q) \oplus x_1. \end{array}\right.$$

Заметим, что определенная таким образом функция $\varphi(x, q)$ является биективной по переменной x, причем:

$$\varphi^{-1}(y, q) = \varphi^{-1}((y_1, y_2), q) = (f(y_1, q) \oplus y_2, y_1),$$

т.е.:

$$\left[\begin{array}{l} x_1 = f(y_1, q) \oplus y_2, \\ x_2 = y_1. \end{array}\right.$$

Функция $f(x_2, q)$ называется *функцией усложнения* шифра Фейстеля. Ценность данного преобразования заключается в том, что даже если $f(x_2, q)$ не является обратимой по x_2, преобразование $\varphi(x, q)$ все равно обратимо по x.

Заметим, что при любом $q \in Q$ подстановку, обратную к подстановке (8.3) по переменной x, можно записать в таком виде:

$$\varphi_q^{-1} = T^m \circ \varphi_q \circ T^m, \tag{8.4}$$

где T^m — циклический сдвиг на m позиций влево и $n = 2m$. Действительно:

$$(T^m \circ \varphi_q \circ T^m)(y_1, y_2) = (T^m \circ \varphi_q)(y_2, y_1) = T^m(y_1, f(y_1, q) \oplus y_2) =$$

$$= (\underbrace{f(y_1, q) \oplus y_2}_{x_1}, \underbrace{y_1}_{x_2}) = (x_1, x_2).$$

Рассмотрим условия обратимости шифра Фейстеля.

Теорема 8.2. Если для любого $q' \in Q'$ выполнено равенство:

$$\alpha_{q'} \circ \beta_{q'} = T^m, \tag{8.5}$$

где T^m — циклический сдвиг влево на m позиций и $n = 2m$, то шифр Фейстеля обратим и правило расшифрования отличается от правила зашифрования только тем, что раундовые ключи используются в обратном порядке.

Доказательство. Для начала заметим, что из равенства (8.5) и равенства $T^m \circ T^m = id$, где id — тождественная подстановка, следуют такие равенства для любого $q' \in Q'$:

$$\begin{aligned} T^m \circ \beta_{q'}^{-1} &= \alpha_{q'}, \\ \alpha_{q'}^{-1} \circ T^m &= \beta_{q'}. \end{aligned} \tag{8.6}$$

Исходя из данных равенств и равенства (8.4), имеем:

$$D_k = \alpha_{q'_0}^{-1} \circ \varphi_{q_1}^{-1} \circ \ldots \circ \varphi_{q_r}^{-1} \circ \beta_{q'_{r+1}}^{-1} \overset{(8.4)}{=}$$

$$\overset{(8.4)}{=} \alpha_{q'_0}^{-1} \circ T^m \circ \varphi_{q_1} \circ T^m \circ \ldots \circ T^m \circ \varphi_{q_r} \circ T^m \circ \beta_{q'_{r+1}}^{-1} \overset{T^m \circ T^m = id}{=}$$

$$= \alpha_{q'_0}^{-1} \circ T^m \circ \varphi_{q_1} \circ \ldots \circ \varphi_{q_r} \circ T^m \circ \beta_{q'_{r+1}}^{-1} \overset{(8.6)}{=}$$

$$\overset{(8.6)}{=} \beta_{q'_0} \circ \varphi_{q_1} \circ \ldots \circ \varphi_{q_r} \circ \alpha_{q'_{r+1}}. \qquad \square$$

Например, если $Q' \subseteq V_m$, то в качестве α и β можно рассмотреть такие отображения:

$$\beta(x, q') = x \oplus q',$$

$$\alpha(x, q') = T^m(x \oplus q').$$

8.4. Построение раундовой функции

Для того, чтобы блочный шифр был вычислительно стойким, необходимо, чтобы раундовая функция $\varphi(x, q)$ (либо функция усложнения $f(x_2, q)$ в случае шифра Фейстеля) удовлетворяла ряду условий.

К. Шеннон сформулировал общий принцип построения шифрующих преобразований, суть которого состоит в том, что незначительное изменение открытого текста или ключа приводило бы к существенному изменению результата. Так как обеспечить данное требование в сочетании с простотой реализацией весьма затруднительно, то К. Шеннон предложил реализовать сложные преобразования в виде суперпозиции нескольких простых некоммутирующих преобразований.

Поэтому функции $\varphi(x, q)$ реализуются в виде композиции нескольких простых преобразований, которые называются *слоями раундовой функции* $\varphi(x, q)$. Каждый слой обеспечивает выполнение одного или нескольких необходимых условий, а их композиция дает в совокупности выполнение всего ряда условий.

Рассмотрим необходимый ряд условий, которым должна удовлетворять раундовая функция.

1. Для любого фиксированного $q \in Q$ функция $\varphi_q : V_n \to V_n$ должна быть **подстановкой**. Данное требование следует из требования обратимости правила зашифрования E_k (см. (8.1)).

Из этого же требования, в частности, следует, что все слои раундовой функции также должны быть обратимыми. Например, в качестве таких слоев часто используются подстановки, перестановки, линейные преобразования и т.д.

2. Если функции $\alpha(x, q')$, $\varphi(x, q)$, $\beta(x, q')$ являются линейными относительно x и раундовых ключей (q или q'), то из равенства (8.1) следует, что отображение E_k также будет являться линейной. В этом случае ключ можно вычислить с помощью системы линейных уравнений. Поэтому раундовая функция должна быть **нелинейной**.

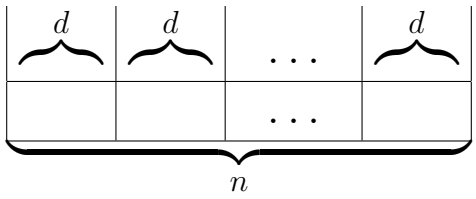
Это условие обеспечивается за счет реализации нелинейного слоя. В качестве такого слоя может выступать нелинейная подстановка $s : V_n \to V_n$.

Так как подстановка, в общем случае, задается с помощью таблицы, то табличная подстановка s должна содержать ровно 2^n столбцов:

$$\begin{pmatrix} 00\ldots00 & 00\ldots01 & \ldots & 11\ldots11 \\ s(00\ldots00) & s(00\ldots01) & \ldots & s(11\ldots11) \end{pmatrix}.$$

При больших n это крайне трудно реализовать. В этом случае поступают следующим образом. Разобьем n-элементный блок

на непересекающиеся подблоки равной длины d :



При этом $n = u \cdot d$, где u, d — целые числа.

Зададим нелинейные биективные преобразования:

$$s_1, s_2, \ldots, s_u : V_d \to V_d.$$

Каждое из данных преобразований обрабатывает лишь часть входного блока. Пусть $x = (x^1, \ldots, x^n) \in V_n$. Тогда:

$$s(x) = s(x^1, \ldots, x^n) =$$

$$= s(x^1, \ldots, x^d,\ x^{d+1}, \ldots,\ x^{2d}, \ldots, x^{(u-1)d+1}, \ldots, x^n) =$$

$$= (s_1(x^1, \ldots, x^d),\ s_2(x^{d+1}, \ldots, x^{2d}), \ldots,\ s_u(x^{(u-1)d+1}, \ldots, x^n)).$$

Преобразования $s_1, \ldots, s_u$ называются *s-боксами* раундовой функции. Каждое из таких преобразований задается в виде таблицы с 2^d столбцами.

3. Каждый s-бокс должен удовлетворять критерию типа **лавинного эффекта**, т.е. изменение одного бита входного набора произвольного s-бокса приводит к изменению нескольких (в среднем не менее двух) битов его выходного набора.

4. Раундовая функция должна содержать **перемешивающие слои**. При этом перемешивающие слои должны удовлетворять условию, что на каждом цикле (начиная со второго) совокупность входных битов каждого s-бокса зависит от выходных битов нескольких s-боксов предыдущего цикла.

В качестве таких слоев выступают перестановки. В простейшем случае это может быть циклическая перестановка.

8.5. Входное и выходное отображения

Входное и выходное отображения нужны для обеспечения обратимости шифра (см. теоремы 8.1 и 8.2) и для усложнения правила зашифрования E_k. Данные отображения в некоторых

итеративных блочных шифрах являются тождественными подстановками id, не зависящими от производного ключа $q' \in Q'$:

$$\alpha(x, q') = x, \quad \beta(x, q') = x.$$

Например, в шифре «Магма» из ГОСТ Р 34.12-2015 для любого $q' \in Q'$ подстановка $\alpha_{q'} = id$, а подстановка $\beta_{q'} = T^m$, где $n = 2m$.

В большинстве случаев множество производных ключей Q' совпадает с V_n, поэтому:

$$\alpha, \beta : V_n \times V_n \to V_n.$$

В качестве отображений α и β можно использовать операции побитового сложения по модулю 2:

$$\alpha(x, q'_0) = x \oplus q'_0, \quad \beta(x, q'_{r+1}) = x \oplus q'_{r+1}.$$

Данные операции получили название *отбеливание текста*, а ключи q'_0 и q'_{r+1} — *ключи отбеливания*. При этом заметим, что:

$$\alpha^{-1}_{q'}(x) = \alpha_{q'}(x), \quad \beta^{-1}_{q'}(x) = \beta_{q'}(x).$$

Отбеливание улучшает криптографические свойства шифра.

8.6. Слабые ключи итеративного блочного шифра

Для начала заметим, что стойкость итеративного блочного шифра тем выше, чем ближе свойства подстановки E_k к свойствам случайных подстановок. Поэтому раундовые ключи $q_1, \ldots, q_r$ должны выбираться случайно из множества Q. Так как мощность множества Q много больше значения r, то из этого следует, что среди раундовых ключей $q_1, \ldots, q_r$ не должно быть одинаковых.

Пусть $k \in K$. Ключ k итеративного r-раундового блочного шифра назовем *μ-слабым*, $1 \leqslant \mu < r$, если набор раундовых ключей $q_1, \ldots, q_r$ содержит ровно μ различных элементов. *Слабым ключом* назовем 1-слабый ключ, т.е. когда $q_1 = \ldots = q_r$.

Теорема 8.3. Пусть k — слабый ключ итеративного r-раундового блочного шифра и $q_1 = \ldots = q_r = q$. Пусть также d —

порядок раундовой функции φ_q (т.е. $\varphi_q^d = id_{V_n}$ и ни для какого натурального числа, меньшего d, это равенство не выполняется) и τ — остаток от деления r на d. Тогда данный блочный шифр на ключе k является τ-раундовым.

Доказательство. Так как $q_1 = \ldots = q_r = q$, то из равенства (8.1) следует, что:

$$E_k = \beta_{q'_{r+1}} \circ \varphi_q^r \circ \alpha_{q'_0}.$$

Из равенств $\varphi_q^d = id_{V_n}$ и $r = ud + \tau$ следует такое равенство:

$$\varphi_q^r = \varphi_q^\tau.$$

Поэтому:

$$E_k = \beta_{q'_{r+1}} \circ \varphi_q^\tau \circ \alpha_{q'_0}.$$

Таким образом, данный блочный шифр на ключе k является τ-раундовым. □

Следствие 8.1. Пусть выполнены условия теоремы 8.3, причем $\tau = 0$ (т.е. d делит нацело r) и подстановки $\alpha_{q'_0}$ и $\beta_{q'_{r+1}}$ являются взаимно обратными. Тогда $E_k = id_{V_n}$.

Теорема 8.4. Пусть k — слабый ключ итеративного $2r$-раундового шифра Фейстеля и $q_1 = \ldots = q_r = q$, $q'_0 = q'_{r+1} = q'$. Пусть также выполнено равенство $\alpha_{q'} \circ \beta_{q'} = T^m$, где $n = 2m$. Тогда будут выполнены следующие условия.

1. Подстановка E_k является инволюцией.

2. Обозначим через G множество всех таких элементов в V_n, что $E_k(x) = x$ для любого $x \in G$. Тогда мощность множества G равна 2^m, где $n = 2m$, т.е.:

$$|G| = \sqrt{|V_n|}.$$

Доказательство. 1. Пусть $q_1 = \ldots = q_r = q$ и $q'_0 = q'_{r+1} = q'$. Тогда из равенства (8.1) следует, что:

$$E_k = \beta_{q'} \circ \varphi_q^{2r} \circ \alpha_{q'}.$$

Из теоремы 8.2 следует, что:

$$E_k^{-1} = D_k = \beta_{q'} \circ \varphi_q^{2r} \circ \alpha_{q'}.$$

Следовательно, подстановка E_k является инволюцией.

2. Так как $\alpha_{q'} \circ \beta_{q'} = T^m$, то правило зашифрования E_k можно записать таким образом:

$$E_k = \alpha_{q'}^{-1} \circ T^m \circ \varphi_q^{2r} \circ \alpha_{q'}.$$

Далее, пусть $\alpha_{q'}(x_1, x_2) = (y_1, y_2)$, где $x_1, x_2, y_1, y_2 \in V_m$. Тогда равенство:

$$E_k(x_1, x_2) = (x_1, x_2)$$

будет выполнено тогда и только тогда, когда выполнено такое равенство:

$$T^m \circ \varphi_q^{2r}(y_1, y_2) = (y_1, y_2).$$

Последнее равенство равносильно такому равенству:

$$T^m \circ \varphi_q^{r}(y_1, y_2) = \varphi^{-r}(y_1, y_2). \tag{8.7}$$

Учитывая, что:

$$\varphi_q^{-r} = T^m \circ \varphi_q^r \circ T^m$$

(см. (8.4)), получаем, что равенство (8.7) равносильно следующему равенству:

$$(y_1, y_2) = T^m(y_1, y_2). \tag{8.8}$$

Так как $T^m(y_1, y_2) = (y_2, y_1)$, то равенство (8.8) возможно тогда и только тогда, когда $y_1 = y_2$.

Таким образом, равенство $E_k(x_1, x_2) = (x_1, x_2)$ равносильно равенству $y_1 = y_2$.

Обозначим через G множество всех таких элементов (x_1, x_2) из V_n, обладающих следующим свойством: элемент (x_1, x_2) принадлежит множеству G тогда и только тогда, когда найдется такой элемент $y \in V_m$, что

$$\alpha_{q'}(x_1, x_2) = (y, y).$$

Так как $\alpha_{q'}$ является подстановкой, то $|G| = 2^m$. □

8.7. Режимы использования блочных шифров

Режим электронной кодовой книги (ECB)

Данный режим соответствует шифру простой замены блоков открытого текста в алфавите V_n. Так как $|V_n| = 2^n$, то правилу зашифрования E_k будет соответствовать подстановка степени 2^n, в соответствии с которой каждый блок открытого текста заменяется блоком шифрованного текста. При этом если $x_1x_2 \dots x_l$ — некоторый открытый текст, разбитый на n-битовые блоки x_1, x_2,. . . ,x_l, а $y_1y_2 \dots y_l$ — соответствующий шифрованный текст, то процесс зашифрования и расшифрования выглядит следующим образом:

$$y_1y_2 \dots y_l = E_k(x_1)E_k(x_2) \dots E_k(x_1),$$

$$x_1x_2 \dots x_l = D_k(y_1)D_k(y_2) \dots D_k(y_1).$$

Отметим следующие особенности данного режима.

1. Так как блоки x_1, x_2,. . . ,x_l открытого текста шифруются независимо друг от друга, то их можно шифровать в произвольном порядке.

2. Если в некотором блоке y_i шифртекста $y_1y_2 \dots y_l$ изменить значения некоторых бит, то при расшифровании это отразится только на соответствующий блок x_i, а остальные блоки шифртекста расшифруются корректно.

Если же в блок y_i добавить некоторое количество дополнительных бит либо из данного блока изъять некоторое количество бит, не кратное числу n, то полученный шифрованный текст $\widetilde{y}_i y_{i+1} \dots y_l$ будет расшифрован некорректно.

3. При удачно построенной шифрующей подстановке E_k, что является общим свойством известных блочных шифров, искажение одного случайно выбранного бита блока x_i приводит к искажению каждого бита в блоке y_i с вероятностью $1/2$. Таким образом, искажение одного бита в блоке x_i приведет в среднем к $n/2$ искажениям в блоке y_i.

4. Одинаковые блоки открытого текста в процессе шифрования перейдут в одинаковые блоки шифртекста. Этот факт позволяет наблюдать частоты появления отдельных блоков. Поэтому данный режим не используется для шифрования длинных сообщений.

Режим сцепления блоков (CBC)

В данном режиме блоки шифрованного текста y_1, y_2,. . . ,y_l вырабатываются по следующему правилу:

$$y_i = E_k(x_i \oplus y_{i-1}), \quad i = 1, 2, \ldots, l,$$

где y_0 — случайный вектор из V_n, называемый *вектором инициализации*. Расшифрование же происходит по такому правилу:

$$x_i = D_k(y_i) \oplus y_{i-1}, \quad i = 1, 2, \ldots, l.$$

Отметим некоторые особенности данного режима.

1. Наличие начального вектора y_0 затрудняет атаку на шифртекст, основанную на наличие стандартов в начале сообщения.

2. Искажение одного бита в блоке x_i приведет к искажению в среднем половины битов во всех блоках шифртекста, начиная с y_i. Однако при расшифровании открытый текст будет содержать ту же единственную ошибку.

3. Если исказить j-ый бит $(1 \leqslant j \leqslant n)$, в блоке y_i, то это приведет к искажению примерно половины битов в блоке x_i и j-го бита в блоке x_{i+1}. Все остальные блоки будут расшифрованы корректно.

Режим гаммирования с обратной связью по шифртексту (CFB)

В данном режиме вырабатывается блочная гамма $\gamma_1\gamma_2 \ldots \gamma_l$ (все $\gamma_i \in V_n$) по следующему правилу:

$$\gamma_i = E_k(y_{i-1}), \quad i = 1, 2, \ldots, l,$$

где y_0 — вектор инициализации. При этом шифртекст $y_1y_2 \ldots y_l$ получается по такому правилу:

$$y_i = x_i \oplus \gamma_i, \quad i = 1, 2, \ldots, l.$$

Тогда процесс расшифрования можно выразить следующими формулами:

$$\gamma_i = E_k(y_{i-1}),$$

$$x_i = y_i \oplus \gamma_i, \quad i = 1, 2, \ldots, l.$$

Приведем некоторые особенности данного режима.

1. Искажение j-го бита в блоке x_i приведет к искажению j-го бита в блоке y_i и в среднем половины битов во всех остальных блоках $(y_{i+1}, \ldots, y_l)$. При расшифровании же открытый текст будет иметь ту же единственную ошибку.

2. Искажение j-го бита в блоке y_i приведет к искажению j-го бита в блоке x_i и примерно половины битов в блоке x_{i+1}. Все другие блоки будут расшифрованы корректно.

3. Отсутствует проблема неполного последнего блока. Пусть блок x_l в сообщении $x_1 x_2 \ldots x_l$ имеет m бит $(1 \leqslant m \leqslant n)$. Тогда после выработки блока гаммы γ_l $(\gamma_l = E_k(y_{l-1}))$ складывается побитово по модулю 2 только первые m битов блока γ_l с блоком x_l и получается m-битовый блок y_l. Расшифрование происходит аналогичным образом.

Режим гаммирования (OFB)

В режиме OFB вырабатывается блочная гамма $\gamma_1 \gamma_2 \ldots \gamma_l$ по такому правилу:

$$\gamma_i = E_k(\gamma_{i-1}), \quad i = 1, 2, \ldots, l,$$

где γ_0 — вектор инициализации. Блоки шифртекста получаются по следующему правилу:

$$y_i = x_i \oplus \gamma_i, \quad i = 1, 2, \ldots, l.$$

Приведем также некоторые особенности данного режима.

1. Искажение j-го бита в блоке x_i приведет только к искажению j-го бита в блоке y_i.

2. Искажение j-го бита в блоке y_i приведет только к искажению j-го бита в блоке x_i.

3. Отсутствует проблема неполного последнего блока.

8.8. Стандарт симметричного блочного шифрования ГОСТ Р 34.12-2015

В стандарте ГОСТ Р 34.12-2015 [17, 18] приведено описание двух базовых блочных шифров с длинами блоков $n = 128$ бит и

$n = 64$ бит и длинами ключей $k = 256$ бит. На шифр с длиной блока $n = 128$ бит можно ссылаться как на блочный шифр «Кузнечик», на шифр с длиной блока $n = 64$ бит — «Магма».

Шифр «Магма»

Данный алгоритм реализует итеративный 32-раундовый обратимый блочный шифр Фейстеля. Размер входного блока — 64 бит. Размер ключа k — 256 бит.

Запишем ключ $k \in V_{256}$ в таком виде: $k = (k_1, k_2, \ldots, k_8)$, где $k_i \in V_{32}$, $i = 1, \ldots, 8$. Раундовые ключи $q_1, \ldots, q_{32}$ принадлежат пространству V_{32} и выбираются из двоичных векторов $k_1, \ldots, k_8$ следующим образом:

$$q_i = \begin{cases} k_{(i-1) \bmod 8 + 1}, & 1 \leqslant i \leqslant 24, \\ k_{8 - (i-1) \bmod 8}, & 25 \leqslant i \leqslant 32, \end{cases}$$

т.е.:

$$(q_1, q_2, \ldots, q_8) = (k_1, k_2, \ldots, k_8),$$
$$(q_9, q_{10}, \ldots, q_{16}) = (k_1, k_2, \ldots, k_8),$$
$$(q_{17}, q_{18}, \ldots, q_{24}) = (k_1, k_2, \ldots, k_8),$$
$$(q_{25}, q_{26}, \ldots, q_{32}) = (k_8, k_7, \ldots, k_1).$$

Входное преобразование α является тождественным, а выходное преобразование $\beta = T^{32}$.

Функция усложнения $f(x_2, q)$ имеет следующие слои:

1. Подмешивание 32-битового раундового ключа путем суммирования по модулю 2^{32} :

$$x_2 + q \pmod{2^{32}}.$$

2. Нелинейная подстановка с помощью s-боксов $s_1, \ldots, s_8$: $V_4 \to V_4$, при этом каждая подстановка s_i представляет собой фиксированную перестановку чисел $\{0, 1, \ldots, 15\}$:

$$\begin{pmatrix} 0 & 1 & \ldots & 15 \\ s_i(0) & s_i(1) & \ldots & s_i(15) \end{pmatrix}.$$

В стандарте s-боксы определены следующим образом:

$s_1 = (1, 7, 14, 13, 0, 5, 8, 3, 4, 15, 10, 6, 9, 12, 11, 2),$
$s_2 = (8, 14, 2, 5, 6, 9, 1, 12, 15, 4, 11, 0, 13, 10, 3, 7),$
$s_3 = (5, 13, 15, 6, 9, 2, 12, 10, 11, 7, 8, 1, 4, 3, 14, 0),$
$s_4 = (7, 15, 5, 10, 8, 1, 6, 13, 0, 9, 3, 14, 11, 4, 2, 12),$
$s_5 = (12, 8, 2, 1, 13, 4, 15, 6, 7, 0, 10, 5, 3, 14, 9, 11),$
$s_6 = (11, 3, 5, 8, 2, 15, 10, 13, 14, 1, 7, 4, 12, 9, 6, 0),$
$s_7 = (6, 8, 2, 3, 9, 10, 5, 12, 1, 14, 4, 7, 11, 13, 0, 15),$
$s_8 = (12, 4, 6, 2, 10, 5, 11, 9, 14, 8, 13, 7, 0, 3, 15, 1),$

$$s(x) = s(x_1, \ldots, x_8) = (s_1(x_1), \ldots, s_8(x_8)),$$
$$x \in V_{32},\ x_i \in V_4,\ i = 1, \ldots, 8.$$

3. Перемешивающий слой, реализованный в виде циклического сдвига 32-битового вектора на 11 бит влево (T^{11}).

Алгоритм шифрования ГОСТ Р 34.12-2015 «Магма»

Пусть на входе имеется 64-битовый блок открытого текста (x_1, x_2), где $x_1, x_2 \in V_{32}$.

Инициализируем начальное значение счетчика $i : i := 1$.

1. Вычисляем значение $y_1 : y_1 := x_2$.

2. Вычисляем значение раундового ключа q :

$$j := \begin{cases} (i-1) \bmod 8 \ + \ 1, & \text{если } 1 \leqslant i \leqslant 24, \\ 8 \ - \ (i-1) \bmod 8, & \text{если } 25 \leqslant i \leqslant 32, \end{cases}$$

$$q := k_j.$$

3. Вычисляем значение $f(x_2, q)$:

$$x_2 := x_2 + q \pmod{2^{32}},$$

$$x_2 := s(x_2),$$
$$f := T^{11}(x_2).$$

4. Вычисляем значение $y_2 = f(x_2, q) \oplus x_1$:

$$y_2 := f \oplus x_1.$$

5. Увеличиваем счетчик: $i := i + 1$.

6. Если $i \leqslant 32$, то присваиваем $x_1 := y_1$, $x_2 := y_2$ и переходим к шагу 1. В противном случае к блоку (y_1, y_2) применяем выходное отображение $\beta = T^{32}$:

$$buf := y_1, \quad y_1 := y_2, \quad y_2 := buf$$

и завершаем преобразования.

На выходе получаем 64-битовый блок шифртекста (y_1, y_2).

Поскольку шифр «Магма» является обратимым, то алгоритм расшифрования блока данных остается прежним, за исключением того, что раундовые ключи располагаются в обратном порядке.

Шифр «Кузнечик»

В данном шифре используются следующие преобразования.

X-преобразование. На вход функции X подаются две последовательности длиной 128 бит каждая, выходом функции является XOR (сложение по модулю два) этих последовательностей:

$$X[k] : V_{128} \to V_{128}, \quad X[k](a) = k \oplus a, \ k, a \in V_{128}.$$

Нелинейное преобразование S. Функция S является подстановкой. Каждый байт из 128-битной входной последовательности заменяется соответствующим байтом из таблицы подстановок π :

$$S : V_{128} \to V_{128}, \quad b = S(a) = S(a_{15}|| \ldots ||a_0) = \pi(a_{15})|| \ldots ||\pi(a_0),$$

где $a = a_{15}|| \ldots ||a_0 \in V_{128}$, $a_i \in V_8$, $i = 0, \ldots 15$, $||$ — операция конкатенации, подстановка $\pi : V_8 \to V_8$ является константой, определенной в стандарте следующим образом:

π = (252, 238, 221, 17, 207, 110, 49, 22, 251, 196, 250, 218, 35, 197, 4, 77, 233, 119, 240, 219, 147, 46, 153, 186, 23, 54, 241. 187, 20, 205, 95, 193, 249, 24, 101, 90, 226, 92, 239, 33, 129, 28, 60, 66, 139, 1, 142, 79, 5, 132, 2, 174, 227, 106, 143, 160, 6, 11, 237, 152, 127, 212, 211, 31, 235, 52, 44, 81, 234, 200, 72, 171, 242, 42, 104, 162, 253, 58, 206, 204, 181, 112, 14, 86, 8, 12, 118, 18, 191, 114, 19, 71,

156, 183, 93, 135, 21, 161, 150, 41, 16, 123, 154, 199, 243, 145, 120, 111, 157, 158, 178, 177, 50, 117, 25, 61, 255, 53, 138, 126, 109, 84, 198, 128, 195, 189, 13, 87, 223, 245, 36, 169, 62, 168, 67, 201, 215, 121, 214, 246, 124, 34, 185, 3, 224, 15, 236, 222, 122, 148, 176, 188, 220, 232, 40, 80, 78, 51, 10, 74, 167, 151, 96, 115, 30, 0, 98, 68, 26, 184, 56, 130, 100, 159, 38, 65, 173, 69, 70, 146, 39, 94, 85, 47, 140, 163, 165, 125, 105, 213, 149, 59, 7, 88, 179, 64, 134, 172, 29, 247, 48, 55, 107, 228, 136, 217, 231, 137, 225, 27, 131, 73, 76, 63, 248, 254, 141, 83, 170, 144, 202, 216, 133, 97, 32, 113, 103, 164, 45, 43, 9, 91, 203, 155, 37, 208, 190, 229, 108, 82, 89, 166, 116, 210, 230, 244, 180, 192, 209, 102, 175, 194, 57, 75, 99, 182).

Заметим, что преобразование, обратное к S, определяется так:

$$S^{-1}(b) = S^{-1}(b_{15}||\dots||b_0) = \pi^{-1}(b_{15})||\dots||\pi^{-1}(b_0).$$

Линейное преобразование L. В стандарте используется конечное поле $GF(2^8)$ с порождающим многочленом:

$$p(x) = Irr(\alpha, \mathbb{Z}_2, x) = x^8 + x^7 + x^6 + x + 1 \in \mathbb{Z}_2[x],$$

причем:

$$GF(2^8) = \langle 1, \alpha, \dots, \alpha^7 \rangle_{\mathbb{Z}_2} \cong \mathbb{Z}_2[x]/\big(p(x)\big).$$

Элементы поля $GF(2^8)$ можно представлять целыми числами в диапазоне от 0 до $2^8 - 1$ следующим образом. Элементу $\sum_{i=0}^{7} x_i\alpha^i \in GF(2^8)$ однозначно соответствует двоичный вектор $(x_7, \dots, x_0) \in V_8$, которому, в свою очередь, однозначно соответствует число $\sum_{i=0}^{7} x_i 2^i$. Поэтому элементу $\sum_{i=0}^{7} x_i\alpha^i$ сопоставим число $\sum_{i=0}^{7} x_i 2^i$. Пусть $\triangle : V_8 \to GF(2^8)$ — биективное отображение, определенное правилом $\triangle(x_7, \dots, x_0) = \sum_{i=0}^{7} x_i\alpha^i$. Обозначим $\nabla = \triangle^{-1}$.

Определим линейное отображение $l : V_8^{16} \to V_8$ следующим образом:

$$l(a_{15},\dots,a_0)=\nabla\big($$

$$\begin{array}{rcrcrcrcr}
 & & 148\cdot\triangle(a_{15}) & + & 32\cdot\triangle(a_{14}) & + & 133\cdot\triangle(a_{13}) & + & 16\cdot\triangle(a_{12}) \\
+ & & 194\cdot\triangle(a_{11}) & + & 192\cdot\triangle(a_{10}) & + & 1\cdot\triangle(a_9) & + & 251\cdot\triangle(a_8) \\
+ & & 1\cdot\triangle(a_7) & + & 192\cdot\triangle(a_6) & + & 194\cdot\triangle(a_5) & + & 16\cdot\triangle(a_4) \\
+ & & 133\cdot\triangle(a_3) & + & 32\cdot\triangle(a_2) & + & 148\cdot\triangle(a_1) & + & 1\cdot\triangle(a_0)\big),
\end{array}$$

где $a_i \in V_8$, $i = 0,\dots,15$, операции сложения и умножения осуществляются в поле $GF(2^8)$, а константы являются элементами поля в указанном ранее смысле. Например, число $148 = (10010100)_2$ соответствует элементу $\alpha^7+\alpha^4+\alpha^2 \in GF(2^8)$.

Нетрудно видеть, что:

$$l(a_{15},\dots,a_1,l(a_{15},\dots,a_1,a_0))=a_0. \tag{8.9}$$

Также определим отображение $R : V_{128} \to V_{128}$ следующим образом:

$$b=R(a)=R(a_{15}||\dots||a_0)=l(a_{15},\dots,a_0)||a_{15}||\dots||a_1,$$

где $a=a_{15}||\dots||a_0 \in V_{128}$, $a_i \in V_8$, $i=0,\dots,15$. Учитывая (8.9), преобразование, обратное к R, определяется так:

$$R^{-1}(b)=R^{-1}(b_{15}||\dots||b_0)=b_{14}||b_{13}||\dots||b_0||l(b_{14},b_{13},\dots,b_0,b_{15}).$$

Обозначим $L(a)=R^{16}(a)=(\underbrace{R\circ\dots\circ R}_{16})(a)$, $a \in V_{128}$. При этом

$$L^{-1}(b)=(R^{-1})^{16}(b)=(\underbrace{R^{-1}\circ\dots\circ R^{-1}}_{16})(b).$$

Алгоритм развертывания ключа. Рассмотрим теперь процедуру генерации раундовых ключей из первичного ключа. Первые два получаются разбиением первичного ключа пополам. Далее для выработки очередной пары раундовых ключей используется 8 итераций сети Фейстеля, где, в свою очередь, в качестве раундовых ключей используются раундовые константы $C_i \in V_{128}$, $i = 1,\dots,32$, которые определены следующим образом:

$$C_i=L(i), \quad i=1,\dots,32.$$

Рассмотрим отображение $F[k] : V_{128} \times V_{128} \to V_{128} \times V_{128}$, которое определяется следующим образом:

$$F[k](a_1, a_0) = \big((L \circ S \circ X[k])(a_1) \oplus a_0, a_1\big), \quad k, a_1, a_0 \in V_{128}.$$

Раундовые ключи $q_i \in V_{128}$, $i = 1, \ldots, 10$, вырабатываются на основе первичного ключа $k = k_1||k_0 \in V_{256}$, $k_0, k_1 \in V_{128}$, и определяются равенствами:

$$q_1 = k_1, \quad q_2 = k_0,$$

$$(q_3, q_4) = \big(F[C_8] \circ \ldots \circ F[C_2] \circ F[C_1]\big)(q_1, q_2),$$
$$(q_5, q_6) = \big(F[C_{16}] \circ \ldots \circ F[C_{10}] \circ F[C_9]\big)(q_3, q_4),$$
$$(q_7, q_8) = \big(F[C_{24}] \circ \ldots \circ F[C_{18}] \circ F[C_{17}]\big)(q_5, q_6),$$
$$(q_9, q_{10}) = \big(F[C_{32}] \circ \ldots \circ F[C_{26}] \circ F[C_{25}]\big)(q_7, q_8),$$

Последние четыре строчки можно записать в виде формулы:

$$(q_{2i+1}, q_{2i+2}) = \big(F[C_{8(i-1)+8}] \circ \ldots \circ F[C_{8(i-1)+1}]\big)(q_{2i-1}, q_{2i}).$$

$$i = 1, 2, 3, 4.$$

Шифрование и расшифрование. Раундовой функцией для шифра «Кузнечик» является отображение:

$$\varphi_q(a) = \varphi(a, q) : V_{128} \times V_{128} \to V_{128},$$

которое определяется следующим образом:

$$\varphi_q(a) = \big(L \circ S \circ X[q]\big)(a).$$

Правила зашифрования и расшифрования для шифра «Кузнечик» определяются так:

$$b = E_k(a) = \big(X[q_{10}] \circ \varphi_{q_9} \circ \ldots \circ \varphi_{q_1}\big)(a),$$
$$a = D_k(b) = \big(\varphi_{q_1}^{-1} \circ \ldots \circ \varphi_{q_9}^{-1} \circ X[q_{10}]\big)(b),$$

где:

$$\varphi_q^{-1} = X[q] \circ S^{-1} \circ L^{-1}, \quad k \in V_{256},\ a, b \in V_{128}.$$

8.9. Шифр AES (Rijndael)

Advanced Encryption Standard (AES) [51], также известный как Rijndael — симметричный алгоритм блочного шифрования, принятый в качестве стандарта шифрования правительством США по результатам конкурса AES. Шифр AES характеризуется размером блока 128 бит, длиной ключа 128, 192 или 256 бит и количеством раундов 10, 12 или 14 в зависимости от длины ключа.

Все операции алгоритма AES работают с байтами (8 бит) данных, поэтому, предполагается, что все переменные и ключи приведены к виду последовательности байтов. Исходный текст состоит из 16 байт, обозначаемых x_0, x_1, ..., x_{15}, представляется в виде двумерного массива байт размера 4×4, записанный по столбцам. Шифр является последовательностью итераций, выполняемых над некоторой промежуточной структурой, называемой *состоянием*. Состояние представляется в виде квадратного массива байтов размера 4×4. После таких преобразований получается шифрованная последовательность байт y_0, y_1, ..., y_{15} :

$$\begin{pmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{pmatrix} \to \begin{pmatrix} s_{0,0} & s_{0,1} & s_{0,2} & s_{0,3} \\ s_{1,0} & s_{1,1} & s_{1,2} & s_{1,3} \\ s_{2,0} & s_{2,1} & s_{2,2} & s_{2,3} \\ s_{3,0} & s_{3,1} & s_{3,2} & s_{3,3} \end{pmatrix} \to \begin{pmatrix} y_0 & y_4 & y_8 & y_{12} \\ y_1 & y_5 & y_9 & y_{13} \\ y_2 & y_6 & y_{10} & y_{14} \\ y_3 & y_7 & y_{11} & y_{15} \end{pmatrix}.$$

Четыре байта каждого столбца массива состояний образуют 32-битное слово, причем номер столбца соответствует номеру слова:

$$w_0 = s_{0,0}s_{1,0}s_{2,0}s_{3,0}, \quad w_1 = s_{0,1}s_{1,1}s_{2,1}s_{3,1},$$

$$w_2 = s_{0,2}s_{1,2}s_{2,2}s_{3,2}, \quad w_3 = s_{0,3}s_{1,3}s_{2,3}s_{3,3}.$$

В шифре AES используется конечное поле $GF(2^8)$ с порождающим многочленом:

$$p(x) = Irr(\alpha, \mathbb{Z}_2, x) = x^8 + x^4 + x^3 + x + 1 \in \mathbb{Z}_2[x],$$

причем:

$$GF(2^8) = \langle 1, \alpha, \ldots, \alpha^7 \rangle_{\mathbb{Z}_2} \cong \mathbb{Z}_2[x]/\big(p(x)\big).$$

Каждый байт данных рассматривается как элемент поля $GF(2^8)$ следующим образом:

$$(x_7 \dots x_1 x_0)_2 \to x_7\alpha^7 + \dots + x_1\alpha + x_0 \in GF(2^8).$$

Сложение байт (элементов поля $GF(2^8)$) представляет собой поразрядное суммирование по модулю 2 — так называемая XOR-суммирование.

Умножение байт происходит с помощью представления их полиномами от α с коэффициентами из $\mathbb{Z}_2$ и перемножения их по обычным алгебраическим правилам. Полученное произведение необходимо привести по модулю многочлена $p(x)$ (т.е. остаток от деления на $p(x)$).

Перемножение многочленов в поле можно упростить, введя произведение многочлена $b(\alpha) = \sum_{i=0}^{7} b_i\alpha^i \in GF(2^8)$ на α :

$$\alpha \cdot (b_7\alpha^7 + b_6\alpha^6 + b_5\alpha^5 + b_4\alpha^4 + b_3\alpha^3 + b_2\alpha^2 + b_1\alpha + b_0) =$$

$$= b_6\alpha^7 + b_5\alpha^6 + b_4\alpha^5 + (b_3 \oplus b_7)\alpha^4 + (b_2 \oplus b_7)\alpha^3 + b_1\alpha^2 + (b_0 \oplus b_7)\alpha + b_7 =$$

$$= \begin{cases} b_6\alpha^7 + b_5\alpha^6 + b_4\alpha^5 + b_3\alpha^4 + b_2\alpha^3 + b_1\alpha^2 + b_0\alpha, & b_7 = 0, \\ b_6\alpha^7 + b_5\alpha^6 + b_4\alpha^5 + (b_3 \oplus 1)\alpha^4 + (b_2 \oplus 1)\alpha^3 + \\ \qquad + b_1\alpha^2 + (b_0 \oplus 1)\alpha + 1, & b_7 = 1. \end{cases}$$

В данном случае использовалось равенство:

$$\alpha^8 = \alpha^4 + \alpha^3 + \alpha + 1.$$

В языке Си умножение байта $b = (b_7, \dots, b_1, b_0)$ на байт (00000010) (который соответствует элементу α) можно записать так:

if $((b >> 7)\ \&\ 1)$
$c = (b \ll 1) \wedge 27;$
else $c = b \ll 1;$

где $27 = (00011011)_2$ соответствует элементу

$$\alpha^4 + \alpha^3 + \alpha + 1 \in GF(2^8).$$

Умножение многочлена $b(\alpha)$ на α^s равносильно применения s-кратной процедуры умножения на α.

Преобразования

Нелинейное преобразование **SubBytes**. Данное преобразование состоит из двух операций:

1. Каждый байт массива состояния заменяется (независимо от других байт) на мультипликативный обратный к нему в поле $GF(2^8)$, причем нулевой байт переходит в нулевой. Для нахождения обратного элемента поля можно использовать обобщенный алгоритм Евклида либо составить таблицу умножения в поле $GF(2^8)$.

2. Над каждым байтом выполняется аффинное преобразование в поле $\mathbb{Z}_2$, задаваемое следующим уравнением:

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Заметим, что последовательности $x_7, \ldots, x_0$ и $y_7, \ldots, y_0$ пронумерованы в естественном следовании бит справа налево. Это аффинное преобразование может быть описано в полиномиальном виде как:

$$b(x) = (x^6 + x^5 + x + 1) + a(x)(x^4 + x^3 + x^2 + x + 1) \bmod (x^8 + 1).$$

Результаты преобразования SubBytes заранее подсчитываются для каждого байта от 0 до 255 и заносятся в таблицу S. Данную таблицу несложно сгенерировать самостоятельно или найти, например, в [51].

Обратным к преобразованию SubBytes будет преобразование, состоящее из обратного аффинного преобразования и взятия мультипликативного обратного в $GF(2^8)$. Результаты преобразования SubBytes^{-1} также заранее подсчитываются для каждого байта от 0 до 255 и заносятся в таблицу S^{-1}.

Преобразование **ShiftRows**. Это преобразование является циклическим сдвигом влево строк массива состояния на раз-

личную величину: i-я строка циклически сдвигается на i байт влево, $i = 0, 1, 2, 3$.

Обратным преобразованием будет циклический сдвиг строк вправо на то же количество позиций.

Преобразование **MixColumns**. В этом преобразовании байты каждого столбца массива состояния рассматриваются как коэффициенты полинома (степени не выше трех) над полем $GF(2^8)$. Преобразование заключается в умножении столбца по модулю $x^4 + 1$ на фиксированный полином:

$$c(x) = 3x^3 + x^2 + x + 2 \in GF(2^8)[x].$$

При этом коэффициент 3 соответствует элементу $\alpha + 1 \in GF(2^8)$, 1 — элементу 1, 2 — элементу α. Полином $c(x)$ взаимно прост с $x^4 + 1$, поэтому умножение обратимо. В матричной форме данное преобразование имеет такой вид:

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Еще раз подчеркнем, что данные преобразования осуществляются над $GF(2^8)$. Сложение 4-байтовых векторов (как и элементов поля $GF(2^8)$) производится путем суммирования $\oplus$ (XOR-суммирования).

Обратное преобразование заключается в умножении каждого столбца по модулю $x^4 + 1$ на полином:

$$d(x) = 11x^3 + 13x^2 + 9x + 14 \in GF(2^8)[x].$$

То же самое в матричной записи выглядит так:

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 14 & 11 & 13 & 9 \\ 9 & 14 & 11 & 13 \\ 13 & 9 & 14 & 11 \\ 11 & 13 & 9 & 14 \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

AddRoundKey — добавление раундового ключа:

$$\mathrm{AddRoundKey}(x, q) = x \oplus q.$$

В каждом раунде используется раундовый ключ W_i (который является частью расширенного ключа), получаемый из секретного ключа K. Размер секретного ключа $l = l(K)$ определяет количество раундов шифра r :

l	128	192	256
r	10	12	14

Расширенный ключ W состоит из блоков по 128 бит (раундовых ключей W_i), количество которых равно количеству раундов плюс 1:

$$W = W_0, W_1, \ldots, W_r.$$

Шифр начинается и заканчивается сложением с ключом.

Алгоритм 8.1 (AES: шифрование блока)**.**
Вход: блок X, расширенный ключ W.
Выход: зашифрованный блок Y.
1. $Y = X \oplus W_0$.
2. for $i = 1, 2, \ldots, r-1$ do
3. $\quad Y = \text{SubBytes}(Y)$.
4. $\quad Y = \text{ShiftRows}(Y)$.
5. $\quad Y = \text{MixColums}(Y)$.
6. $\quad Y = Y \oplus W_i$.
7. $Y = \text{SubBytes}(Y)$.
8. $Y = \text{ShiftRows}(Y)$.
9. $Y = Y \oplus W_r$.
10. return Y.

Расшифрование блока данных проходит путем последовательного применения рассмотренных преобразований в обратном порядке.

Алгоритм 8.2 (AES: расшифрование блока)**.**
Вход: зашифрованный блок Y, расширенный ключ W.
Выход: блок X.
1. $X = Y \oplus W_r$;
2. for $i = r-1, r-2, \ldots, 1$ do
3. $\quad X = \text{ShiftRows}^{-1}(X)$.

4. $X = \text{SubBytes}^{-1}(X)$.
5. $X = Y \oplus W_i$.
6. $X = \text{MixColums}^{-1}(X)$.
7. $X = \text{ShiftRows}^{-1}(X)$.
8. $X = \text{SubBytes}^{-1}(X)$.
9. $X = X \oplus W_0$.
10. return X.

Данный алгоритм можно записать по-другому. Понятно, что операции SubBytes и ShiftRows (как и операции SubBytes^{-1} и ShiftRows^{-1}) можно менять местами. Также верно такое равенство:

$$\text{MixColums}^{-1}(Y \oplus W_i) = \text{MixColums}^{-1}(Y) \oplus \text{MixColums}^{-1}(W_i).$$

Поэтому если обозначить

$$\widetilde{W_i} = \text{MixColums}^{-1}(W_i), \quad i = 1, \ldots, r-1,$$

то алгоритм расшифрования можно записать следующим образом.

Алгоритм 8.3 (AES: расшифрование блока)**.**
Вход: зашифрованный блок Y, расширенный ключ W.
Выход: блок X.
1. $X = Y \oplus W_r$.
2. for $i = r-1, r-2, \ldots, 1$ do
3. $X = \text{SubBytes}^{-1}(X)$.
4. $X = \text{ShiftRows}^{-1}(X)$.
5. $X = \text{MixColums}^{-1}(X)$.
6. $X = Y \oplus \widetilde{W_i}$.
7. $X = \text{SubBytes}^{-1}(X)$.
8. $X = \text{ShiftRows}^{-1}(X)$.
9. $X = X \oplus W_0$.
10. return X.

Алгоритм расширения ключа. Назовем словом последовательность из четырех байт (32 бита). Байты в слове нумеруются от 0 до 3 слева направо. В алгоритмах зашифрования и

расшифрования было удобно делить расширенный ключ W на слова по четыре слова (раундовые ключи W_i). Однако формирование ключа проходит в пословном режиме. Поэтому будем обозначать буквой w с индексом отдельное слово в W, нумеруя слова с нуля. Напомним, что расширенный ключ W состоит из $r+1$ блоков (последовательность из четырех слов — 128 бит), где r — количество раундов в шифре. Поэтому количество слов в W равно $4(r+1)$:

$$W = w_0 w_1 \dots w_{4(r+1)-1}.$$

Обозначим через c число слов в секретном (первичном) ключе K. Ниже приведена зависимость значения c от числа бит l ключа K :

l	128	192	256
	4	6	8

Расширение ключа K происходит по одному из двух вариантов рекуррентного закона (в зависимости от длины ключа K).

Алгоритм 8.4 (AES: формирование расширенного ключа)**.**

Вход: секретный ключ K из c слов.
Выход: расширенный ключ W из $4(r+1)$ слов.

1. В w_0, w_1, …, w_{c-1} записываем все c слов ключа K.
2. for $i = c, c+1, \dots, 4(r+1)-1$ do
3. $\quad t = w_{i-1}$.
4. $\quad$ if $i \equiv 0 \pmod{c}$ then
5. $\quad\quad t = \mathrm{SubWord}(\mathrm{RotWord}(t)) \oplus \mathrm{Rcon}[i \text{ div } c]$
6. $\quad$ else if $c = 8$ and $i \equiv 4 \pmod{c}$ then
7. $\quad\quad t = \mathrm{SubWord}(t)$.
8. $\quad w_i = t \oplus w_{i-c}$.
9. return $w_0 w_1 \dots w_{4(r+1)-1}$.

В данном алгоритме $\mathrm{RotWord}(t)$ — преобразование, осуществляющее циклический сдвиг байт в слове t на одну позицию влево:

$$(t_0, t_1, t_2, t_3) \to (t_1, t_2, t_3, t_0).$$

SubWord(t) — функция, применяющая S-бокс шифра к каждому байту слова t :

$$\text{Subword}(t) = (S(t_0), S(t_1), S(t_2), S(t_3)).$$

Массив раундовых констант (слов) Rcon$[i]$, $i = 1, 2, \ldots, N$ ($N = 10$ при $l = 128$, $N = 8$ при $l = 192$), состоит из слов вида $(RC[i], 0, 0, 0)$, $RC[i] \in GF(2^8)$, которые меняются по рекурсивному закону:

$$RC[1] = 1, \quad RC[i] = 2 \cdot RC[i-1],\ i = 2, 3, \ldots, N.$$

где число 2 соответствует элементу $\alpha \in GF(2^8)$ (алгоритм умножения на α описан выше) и умножение — умножение в поле $GF(2^8)$.

Аналитическая форма записи правил зашифрования и расшифрования. Запишем алгоритм зашифрования блока данных в несколько ином виде.

Разобьем раундовый ключ на байтовые ключи:

$$W = W_0 W_1 \ldots W_r = w_0 w_1 \ldots w_{4r+3} = q_0 q_1 \ldots q_{16r+15},$$

где все $W_i \in V_{128}$, $w_j \in V_{32}$, $q_k \in V_8$, причем:

W_0	W_1	$\ldots$	W_i	$\ldots$	W_r
$w_0 \ldots w_3$	$w_4 \ldots w_7$	$\ldots$	$w_{4i} \ldots w_{4i+3}$	$\ldots$	$w_{4r} \ldots w_{4r+3}$
$q_0 \ldots q_{15}$	$q_{16} \ldots q_{31}$	$\ldots$	$q_{16i} \ldots q_{16i+15}$	$\ldots$	$q_{16i} \ldots q_{16i+15}$

На основе преобразований SubBytes, ShiftRows и MixColums определим следующие отображения:

$$\varphi_{W_i}(X) = \varphi(X, W_i) : V_{128} \times V_{128} \to V_{128},$$

где:

$$\varphi_{W_i}(X) = \varphi_{W_i} \begin{pmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{pmatrix} =$$

$$= \begin{pmatrix} 2 & 3 & 1 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 3 \\ 3 & 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} S(x_0) & S(x_4) & S(x_8) & S(x_{12}) \\ S(x_5) & S(x_9) & S(x_{13}) & S(x_1) \\ S(x_{10}) & S(x_{14}) & S(x_2) & S(x_6) \\ S(x_{15}) & S(x_3) & S(x_7) & S(x_{11}) \end{pmatrix} \oplus$$

$$\oplus \begin{pmatrix} q_{16i} & q_{16i+4} & q_{16i+8} & q_{16i+12} \\ q_{16i+1} & q_{16i+5} & q_{16i+9} & q_{16i+13} \\ q_{16i+2} & q_{16i+6} & q_{16i+10} & q_{16i+14} \\ q_{16i+3} & q_{16i+7} & q_{16i+11} & q_{16i+15} \end{pmatrix}.$$

$$\alpha_{W_i}(X) = \alpha(X, W_i) : V_{128} \times V_{128} \to V_{128},$$

где:

$$\alpha_{W_i}(X) = \alpha_{W_i} \begin{pmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{pmatrix} =$$

$$= \begin{pmatrix} S(x_0) & S(x_4) & S(x_8) & S(x_{12}) \\ S(x_5) & S(x_9) & S(x_{13}) & S(x_1) \\ S(x_{10}) & S(x_{14}) & S(x_2) & S(x_6) \\ S(x_{15}) & S(x_3) & S(x_7) & S(x_{11}) \end{pmatrix} \oplus \begin{pmatrix} q_{16i} & q_{16i+4} & q_{16i+8} & q_{16i+12} \\ q_{16i+1} & q_{16i+5} & q_{16i+9} & q_{16i+13} \\ q_{16i+2} & q_{16i+6} & q_{16i+10} & q_{16i+14} \\ q_{16i+3} & q_{16i+7} & q_{16i+11} & q_{16i+15} \end{pmatrix}.$$

$$A_{W_i}(X) : V_{128} \times V_{128} \to V_{128}, \quad A_{W_i}(X) = X \oplus W_i.$$

Также заметим, что все операции в данных преобразованиях осуществляются над полем $GF(2^8)$.

Правила зашифрования и расшифрования для шифра AES определяются следующим образом:

$$E_K(X) = (\alpha_{W_r} \circ \varphi_{W_{r-1}} \circ \ldots \circ \varphi_{W_2} \circ \varphi_{W_1} \circ A_{W_0})(X),$$

$$D_K(Y) = (\beta_{W_0} \circ \psi_{\widetilde{W}_1} \circ \ldots \circ \psi_{\widetilde{W}_{r-2}} \circ \psi_{\widetilde{W}_{r-1}} \circ A_{W_r})(Y),$$

где:

$$\psi_{\widetilde{W}_i}(X) = \psi(X, \widetilde{W}_i) : V_{128} \times V_{128} \to V_{128},$$

$$\psi_{\widetilde{W}_i}(X) = \psi_{\widetilde{W}_i} \begin{pmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{pmatrix} =$$

$$= \begin{pmatrix} 14 & 11 & 13 & 9 \\ 9 & 14 & 11 & 13 \\ 13 & 9 & 14 & 11 \\ 11 & 13 & 9 & 14 \end{pmatrix} \cdot \begin{pmatrix} S^{-1}(x_0) & S^{-1}(x_4) & S^{-1}(x_8) & S^{-1}(x_{12}) \\ S^{-1}(x_{13}) & S^{-1}(x_1) & S^{-1}(x_5) & S^{-1}(x_9) \\ S^{-1}(x_{10}) & S^{-1}(x_{14}) & S^{-1}(x_2) & S^{-1}(x_6) \\ S^{-1}(x_7) & S^{-1}(x_{11}) & S^{-1}(x_{15}) & S^{-1}(x_3) \end{pmatrix} \oplus$$

$$\oplus \begin{pmatrix} \widetilde{q}_{16i} & \widetilde{q}_{16i+4} & \widetilde{q}_{16i+8} & \widetilde{q}_{16i+12} \\ \widetilde{q}_{16i+1} & \widetilde{q}_{16i+5} & \widetilde{q}_{16i+9} & \widetilde{q}_{16i+13} \\ \widetilde{q}_{16i+2} & \widetilde{q}_{16i+6} & \widetilde{q}_{16i+10} & \widetilde{q}_{16i+14} \\ \widetilde{q}_{16i+3} & \widetilde{q}_{16i+7} & \widetilde{q}_{16i+11} & \widetilde{q}_{16i+15} \end{pmatrix},$$

$$\beta_{W_i}(X) = \beta(X, W_i) : V_{128} \times V_{128} \to V_{128},$$

$$\beta_{W_i}(X) = \beta_{W_i} \begin{pmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{pmatrix} =$$

$$= \begin{pmatrix} S^{-1}(x_0) & S^{-1}(x_4) & S^{-1}(x_8) & S^{-1}(x_{12}) \\ S^{-1}(x_{13}) & S^{-1}(x_1) & S^{-1}(x_5) & S^{-1}(x_9) \\ S^{-1}(x_{10}) & S^{-1}(x_{14}) & S^{-1}(x_2) & S^{-1}(x_6) \\ S^{-1}(x_7) & S^{-1}(x_{11}) & S^{-1}(x_{15}) & S^{-1}(x_3) \end{pmatrix} \oplus$$

$$\oplus \begin{pmatrix} q_{16i} & q_{16i+4} & q_{16i+8} & q_{16i+12} \\ q_{16i+1} & q_{16i+5} & q_{16i+9} & q_{16i+13} \\ q_{16i+2} & q_{16i+6} & q_{16i+10} & q_{16i+14} \\ q_{16i+3} & q_{16i+7} & q_{16i+11} & q_{16i+15} \end{pmatrix}.$$

8.10. Основы криптоанализа

8.10.1. Криптоатаки

Главная цель шифра — надежная защита скрываемой информации. Криптоанализ шифра может проводиться исходя из той или иной исходной информации, располагаемой противником. Напомним, что в определение шифра входят такие параметры как ключ шифрования k, алгоритм шифрования E_k на ключе k, алгоритм расшифрования D_k на ключе k, открытый текст x, шифртекст $y = E_k(x)$.

В криптографии с секретным ключом рассматривают следующие атаки, целью которых определить ключ шифрования k.

- *Атака на основе шифртекста.* Криптоаналитик располагает (одним или несколькими) шифртекстами $y_1,\ldots,y_m$, полученными из неизвестных открытых текстов $x_1,\ldots,x_m$ на секретном ключе $k : y_i = E_k(x_i)$, $i = 1, \ldots, m$.

- *Атака на основе известного открытого текста.* Криптоаналитик располагает (одной или несколькими) парами $(x_1, y_1),\ldots, (x_m, y_m)$ открытых и соответствующих им шифрованных текстов. Требуется найти ключ $k : y_i = E_k(x_i)$, $i = 1, \ldots, m$.

- *Атака на основе выбранного открытого текста.* Отличается от предыдущей лишь тем, что криптоаналитик имеет возможность выбора открытых текстов $x_1,\ldots,x_m$. Данная атака возможна, например, если криптоаналитик имеет доступ к шифратору передающей стороны.

- *Атака на основе выбранного шифрованного текста.* Отличается от предыдущей лишь тем, что криптоаналитик имеет возможность выбора шифртекстов $y_1,\ldots,y_m$. Данная атака возможна, например, если криптоаналитик имеет доступ к шифратору принимающей стороны.

Атаки на основе выбранных открытых текстов являются наиболее опасными. При этом предполагается, что выполнен так называемый *принцип Кирхгофа* — принцип построения криптографических алгоритмов, согласно которому в секрете держится только определенный набор параметров алгоритма (ключ), а остальные детали могут быть открыты без снижения стойкости алгоритма ниже допустимой величины.

Рассмотрим некоторые универсальные методы криптоанализа, предполагая, что криптоаналитик располагает одной или несколькими парами $(x_1, y_1),\ldots,(x_m, y_m)$, знает алгоритм зашифрования и расшифрования, но ему неизвестен только секретный ключ k.

8.10.2. Метод полного перебора

Рассмотрим уравнение относительно $k \in K$ при известной паре $(x, y) \in X \times Y$:

$$E_k(x) = y. \tag{8.10}$$

Будем для простоты полагать, что для любой пары $(x, y) \in X \times Y$ существует, и притом единственный, ключ $k \in K$ такой, что $E_k(x) = y$. В данном методе криптоанализа последовательно перебираются ключи из K и подставляются в уравнение (8.10) до тех пор, пока оно не станет верным. Будем предполагать, что проверка одного ключа $k \in K$ в уравнении (8.10) это одна операция. Пусть t_k — число таких операций,

необходимых для решения уравнения (8.10) методом перебора. Так как множество K конечно, пронумеруем его некоторым образом: $K = \{k_1, \ldots, k_m\}$. Пусть $P(K)$ — распределение вероятностей на множестве K рассматриваемого шифра Σ_B. Заметим, что t_k является случайной величиной, определенной на множестве K, с областью значений $\{1, 2, \ldots, m\}$ и распределением вероятностей:

$$\begin{array}{cccc} 1 & 2 & \ldots & m \\ P_K(k_1) & P_K(k_2) & \ldots & P_K(k_m) \end{array}$$

Тогда математическое ожидание случайной величиной t_K примет такой вид:

$$M(t_k) = \sum_{i=1}^{m} i \cdot P_K(k_i).$$

Предположим, что распределение вероятностей $P(K)$ равномерно. Тогда:

$$M(t_k) = \frac{1}{m} \sum_{i=1}^{m} i = \frac{m+1}{2} = \frac{|K|+1}{2},$$

т.е. в данном случае в среднем придется перебрать примерно $|K|/2$ ключей для решения уравнения (8.10) данным методом. Например, средняя трудоемкость полного перебора в шифре «Магма» из ГОСТ Р 34.12-2015 в среднем займет приблизительно 2^{255} операций (подстановок ключей).

Алгоритмы полного перебора можно распараллеливать, что позволяет значительно ускорить нахождение ключа.

Пусть имеются N процессоров (например, суперкомпьютер или N независимо работающих компьютеров). Разобьем множество K на N непересекающихся подмножеств $K_1, \ldots, K_N$ и i-й процессор будет перебирать ключи из множества K_i, где $i = 1, \ldots, N$. При нахождении ключа одним из процессоров, он сигнализирует другим процессорам о прекращении работы. Если мощности множеств $K_1, \ldots, K_N$ примерно одинаковы, то для нахождения ключа потребуется в N раз меньше времени: $t_k \approx \dfrac{|K|}{2N}$. Например, если некоторый шифр Σ_B использует дли-

ну ключа 2^{32} и имеется $2^{16} = 65536$ процессоров (компьютеров), то каждый из данных процессоров в среднем проведет $\frac{2^{32}}{2 \cdot 2^{16}} = 2^{15}$ опробований ключей для решения (8.10). Если один процессор на одно опробование ключа тратит 2^{-8} секунды, то для нахождения ключа в среднем потребуется $2^{15} \cdot 2^{-8} = 128$ секунд.

В настоящее время очень широкое распространение получили глобальные сети, поэтому можно задействовать очень много компьютеров через программу-вирус, которая опробует ключи.

В данном методе самой большой сложностью является разбиение множества K. Но можно поступить и следующим образом. Для начала напомним, что дискретная случайная величина ξ имеет *геометрический закон распределения* с параметром p, $0 < p < 1$, если она принимает значения $1, 2, \ldots, n, \ldots$ (областью значений ξ является множество натуральных чисел) с вероятностями $P(\xi = n) = p(1-p)^{n-1}$.

Предложение 8.1. Математическое ожидание случайной величины, имеющей геометрический закон распределения с параметром p, $0 < p < 1$, равно $1/p$.

Доказательство. Учитывая определение математического ожидания случайной величины, имеем:

$$M\xi = \sum_{n=1}^{+\infty} np(1-p)^{n-1}.$$

Рассмотрим функцию:

$$F(t) = p \sum_{n=1}^{+\infty} n(1-p)^{n-1} t^{n-1}.$$

Тогда:

$$\int F(t)dt = p \sum_{n=1}^{+\infty} (1-p)^{n-1} t^n + C =$$

$$= pt \sum_{n=1}^{+\infty} \Big((1-p)t\Big)^{n-1} + C = \frac{pt}{1-(1-p)t} + C.$$

В последнем равенстве использовалось $0 < (1-p)t < 1$. Беря производную от последнего значения, получаем:

$$F(t) = \frac{p\Big(1-(1-p)t\Big) + pt(1-p)}{\Big(1-(1-p)t\Big)^2},$$

$$M\xi = F(1) = \frac{p\Big(1-(1-p)\Big) + p(1-p)}{\Big(1-(1-p)\Big)^2} = \frac{p^2+p-p^2}{p^2} = \frac{1}{p}. \quad \square$$

Пусть процессоры выбирают ключи из K случайным (равновероятным) образом. Тогда вероятность того, что данный процессор выберет не тот ключ равна $1 - \dfrac{1}{|K|}$, а вероятность того, что все N процессоров, работающих синхронно, выберут не те ключи равна:

$$q = \left(1 - \frac{1}{|K|}\right)^N.$$

Вероятность же успеха равна $p = 1 - q$. Таким образом, получаем геометрическое распределение:

$$\begin{array}{cccccc} 1 & 2 & \dots & n & \dots \\ p & qp & \dots & q^{n-1}p & \dots, \end{array}$$

т.е. вероятность того, что на n-ом шаге один из процессоров выберет правильный ключ равна $q^{n-1}p$. В среднем потребуется:

$$\frac{1}{p} = \frac{1}{1 - \left(1 - \frac{1}{|K|}\right)^N}$$

шагов. Так как $|K| \gg N$, то, учитывая приближенное равенство:

$$\left(1 - \frac{1}{|K|}\right)^N \approx 1 - \frac{N}{|K|},$$

в среднем потребуется примерно $\dfrac{|K|}{N}$ шагов (опробований ключей каждым процессором).

Таким образом, при случайном выборе ключей время перебора N процессорами увеличивается в 2 раза, но исчезает проблема разбиения множества. Если же количество процессоров, случайно выбирающих ключи, увеличить в 2 раза, то время перебора будет примерно таким же, что и в предыдущем способе, связанным с разбиением множества K.

8.10.3. Аналитический метод

Пусть множество K можно представить в виде декартова произведения $K = K_1 \times \ldots \times K_m$. Например, для шифра «Магма» из ГОСТ Р 34.12.2015 $K = V_{256}$. Пусть $x = x_1 \ldots x_l \in X$, $y = y_1 \ldots y_l \in Y$ и $y = E_k(x)$. Тогда для некоторых функций $f_1,\ldots,f_l$ имеет место запись:

$$\begin{array}{rcl} y_1 & = & f_1(x_1, \ldots, x_l, k_1, \ldots, k_m) \\ & \ldots & \\ y_l & = & f_l(x_1, \ldots, x_l, k_1, \ldots, k_m) \end{array} \tag{8.11}$$

В данных уравнениях известны f, x, а неизвестен только ключ k. Рассмотрим варианты относительно системы (8.11).

1. Предположим, что система (8.11) может быть приведена к виду:

$$\begin{array}{rcl} g_1(x, y) & = & h_1(x, y, k_1) \\ g_2(x, y) & = & h_2(x, y, k_1, k_2) \\ & \ldots & \\ g_m(x, y) & = & h_m(x, y, k_1, k_2, \ldots, k_m) \end{array} \tag{8.12}$$

Тогда первое уравнение системы (8.12) можно решить методом перебора ключа $k_1 \in K_1$, опробуя при этом в среднем $|K_1|/2$ ключей из K_1. После нахождения ключа k_1, перебираем ключи $k_2 \in K_2$ для решения второго уравнения системы (8.12) и т.д. В результате весь ключ $k = (k_1, \ldots, k_m)$ можно найти методом перебора, опробуя при этом в среднем $(|K_1| + \ldots + |K_m|)/2$ ключей. Например, если $|K_1| = \ldots = |K_m| = \sqrt[m]{|K|}$, то в среднем опробований будет $m\sqrt[m]{|K|}/2$.

2. Предположим, что из (8.11) каким-либо образом можно

выделить линейную подсистему уравнений:

$$\begin{aligned} g_1(x,y) &= a_{11}k_1 + \ldots + a_{1m}k_m \\ &\ldots \\ g_m(x,y) &= a_{m1}k_1 + \ldots + a_{mm}k_m \end{aligned}$$

применяя для этой системы метод Гаусса, оценим его трудоемкость. Для получения первой ступеньки необходимо не более $(m+1)(m-1) < m^2$ операций. Для второй ступеньки — не более $(m-1+1)(m-1-1) < (m-1)^2$ операций и т.д. Всего необходимо произвести не более:

$$\sum_{i=1}^{m} i^2 = \frac{m(m+1)(2m+1)}{6} \approx \frac{m^3}{3}$$

операций. Например, для нахождения 64-битного ключа методом полного перебора необходимо в среднем 2^{63} опробований ключа, но если задача сводится к решению системы линейных уравнений над полем $\mathbb{Z}_2$, то трудоемкость составит не более $\frac{64^3}{3} < 2^{17}$ операций.

8.10.4. Метод встречи посередине

Пусть шифр $\Sigma_A = (X_1, K_1 \times K_2, Y_2, E, D)$ является произведением шифров:

$$\Sigma_{A_1} = (X_1, K_1, Y_1, E^{(1)}, D^{(1)}), \quad \Sigma_{A_2} = (X_2, K_2, Y_2, E^{(2)}, D^{(2)}),$$

где $Y_1 = X_2$. Пусть $y = E^{(2)}_{k_2}(E^{(1)}_{k_1}(x))$, $(x,y) \in X_1 \times Y_2$. Трудоемкость полного перебора равна $\frac{|K_1||K_2|}{2}$. Попробуем уменьшить данную сложность за счет увеличения используемой памяти.

Преобразуем уравнение $E^{(2)}_{k_2}(E^{(1)}_{k_1}(x)) = y$ к виду $E^{(1)}_{k_1}(x) = D^{(2)}_{k_2}(y)$. Для всех $k_1^{(i)} \in K_1$, $k_1^{(i)} = 1, \ldots, |K_1|$, построим таблицу:

$$\begin{aligned} z_1 &= E^{(1)}_{k_1^{(1)}}(x), \\ &\ldots \\ z_{|K_1|} &= E^{(1)}_{k_1^{(|K_1|)}}(x). \end{aligned}$$

Построим еще одну таблицу для всех $k_2^{(j)} \in K_2$, $j = 1, \ldots, |K_2|$:

$$\widetilde{z}_1 = D^{(2)}_{k_2^{(1)}}(y),$$
$$\ldots$$
$$\widetilde{z}_{|K_2|} = D^{(2)}_{k_2^{(|K_2|)}}(y).$$

Объединим две данные таблицы и проведем упорядочивание в полученной таблице в соответствие с некоторым порядком на множестве $Y_1 = X_2$. Известно, что сложность упорядочивания оценивается величиной:

$$(|K_1| + |K_2|) \ln(|K_1| + |K_2|).$$

Итак, число затраченных операций составляет:

$$(|K_1| + |K_2|) + (|K_1| + |K_2|) \ln(|K_1| + |K_2|).$$

Также в данном методе можно поступить несколько иначе. Можно организовать память таким образом, что значения элементов множества $Y_1 = X_2$ являются адресами данной памяти. Сначала для каждого ключа $k_1 \in K_1$ по адресу $E^{(1)}_{k_1}(x)$ запишем ключ k_1. Затем для каждого $k_2 \in K_2$ по адресу $D^{(2)}_{k_2}(y)$ производится проверка значений. Если по данному адресу хранится ключ k_1, то образуется допустимая пара (k_1, k_2). При этом заметим, что по адресу $D^{(2)}_{k_2}(y)$ могут находиться несколько ключей из множества K_1, поэтому память необходимо должным образом организовать. Таким образом, в данном случае требуется уже $|K_1| + |K_2|$ опробований и столько же операций обращения к памяти.

Глава 9. Шифрование с открытым ключом

В симметричной криптографии каждая из переписывающихся сторон должна иметь копию общего секретного ключа, что создает сложнейшую проблему управления ключами. В криптосистемах, о которых пойдет речь в этой главе, используются два ключа: открытый и секретный. Открытый ключ может быть опубликован в общедоступном справочнике наряду с именем пользователя. В результате любой желающий может зашифровать с его помощью свое сообщение и послать закрытую информацию владельцу соответствующего секретного ключа. Расшифровать посланное сообщение сможет только тот, у кого есть секретный ключ.

9.1. Предыстория и основные идеи

Рассмотрим две задачи, решение которых поможет лучше понять идеи и методы криптографии с открытым ключом [41].

Первая задача — хранение паролей в компьютере. Известно, что каждый пользователь в сети имеет свой секретный пароль. При входе в сеть пользователь указывает свое имя (несекретное) и пароль. Проблема состоит в том, что нежелательно хранить пароль на сервере, так как этот пароль могут несанкционированно использовать, например, администратор сети.

Вторая задача связана с появлением радиолокаторов и системы ПВО. При пересечении самолетом границы радиолокатор запрашивает пароль. Если пароль верный, то это свой самолет, в противном случае — чужой. Проблема данной задачи состоит в том, что противник может подслушать верный пароль, так как все переговоры происходят по открытому каналу, и в слу-

чае запроса радиолокатором пароля противник может просто передать перехваченный пароль и будет пропущен.

Данные задачи можно решить с помощью криптографических методов.

Пусть $y = f(x)$ — некоторая функция, определенная на конечном множестве X, для которой существует обратная функция $x = f^{-1}(y)$. Будем говорить, что функция f является *односторонней*, если вычисление значения $f(x)$, $x \in X$, является простой задачей, однако для данного $y \in f(X)$ из области ее значений вычислительно сложно нахождение значения аргумента x, для которого $f(x) = y$.

Пусть p — некоторое простое число и $X = \{0, 1, \ldots, p-1\}$. Зафиксируем некоторое $a \in X$. В качестве примера односторонней функции рассмотрим такую функцию:

$$y = f(x) = a^x \pmod{p},\ x \in X.$$

Обратная функция обозначается:

$$x = f^{-1}(y) = \log_a y \pmod{p}$$

и называется *дискретным логарифмом.*

Даже для очень больших значений числа p (например, 1024-битного) для заданного x легко вычислить значение функции $f(x)$. Однако вычисление обратной задачи очень трудоемко в случае, если число $p-1$ содержит один простой делитель, например, когда $p - 1 = 2q$, где q — простое число.

Напомним бинарный алгоритм быстрого вычисления значения $a^x \pmod{p}$. Обозначим $t = [\log_2 x]$, где [ ] — целая часть числа. Рассмотрим такую последовательность:

$$a,\ a^2,\ a^4,\ a^8, \ldots,\ a^{2^t} \pmod{p}. \tag{9.1}$$

При этом для любого $i = 1, 2, \ldots, t$ выполняется равенство:

$$a^{2^i} \equiv a^{2^{i-1}} \cdot a^{2^{i-1}} \pmod{p},$$

т.е. каждое число рассматриваемой последовательности получается путем умножения предыдущего числа самого на себя по

модулю p. Представим число x (показатель степени) в виде такого разложения:

$$x = x_t 2^t + x_{t-1} 2^{t-1} + \ldots + x_2 2^2 + x_1 2 + x_0,$$

где все $x_i \in \{0, 1\}$, т.е. число x в двоичной системе счисления имеет такую запись:

$$x = (x_t x_{t-1} \ldots x_2 x_1 x_0)_2.$$

Тогда:

$$a^x = a^{x_0} \cdot a^{x_1 \cdot 2} \cdot a^{x_2 \cdot 4} \ldots \cdot a^{x_t \cdot 2^t}.$$

Так как:

$$a^{x_i \cdot 2^i} = \begin{cases} a^{2^i}, & x_i = 1, \\ 1, & x_i = 0, \end{cases}$$

то для нахождения значения a^x необходимо найти значения элементов последовательности (9.1) по модулю p, а затем перемножить те из них, для которых соответствующие значения $x_i = 1$.

Пример 9.1. Вычислим $3^{50} \pmod{11}$. $t = [\log_2 50] = 5$. Вычислим числа ряда (9.1):

$$\begin{array}{cccccc} 3 & 3^2 & 3^4 & 3^8 & 3^{16} & 3^{32} \pmod{11} \\ 3 & 9 & 4 & 5 & 3 & 9 \pmod{11}. \end{array}$$

Далее представим число 50 в двоичной системе счисления:

$$50 = 110010_2.$$

Следовательно:

$$3^{50} \equiv 3^2 \cdot 3^{16} \cdot 3^{32} \equiv 9 \cdot 3 \cdot 9 \equiv 1 \pmod{11}.$$

Замечание 9.1. Если требуется вычислить $a^x \pmod{p}$ при $x \geqslant p - 1$, то вычисления можно упростить, используя теорему Ферма. Так как $a^{p-1} \equiv 1 \pmod{p}$ при $0 < a < p$, то сначала требуется разложить число x по модулю $p - 1$:

$$x = (p - 1)q + r, \quad 0 \leqslant r < p - 1,$$

а затем вычислить $a^r \pmod{p}$, так как $a^x \equiv a^r \pmod{p}$. Поэтому в примере 9.1 $3^{50} = (3^{10})^5 \equiv (1)^5 = 1 \pmod{11}$.

В общем случае получается такое предложение.

Предложение 9.1. Чтобы вычислить значение $a^x \pmod m$, где a, x и m — некоторые натуральные числа, достаточно выполнить не более $2 \cdot \log_2 x$ умножений.

Доказательство. Для вычислений чисел последовательности (9.1) требуется $t = [\log_2 x]$ умножений. Затем при вычислении $a^x \pmod m$ перемножаются некоторые из чисел последовательности (9.1), т.е. не более t умножений. Осталось заметить, что $[\log_2 x] \leqslant \log_2 x$. □

Рассмотрим решение описанных выше задач с помощью односторонней функции $f(x) = a^x \pmod p$. Пусть числа a и p зафиксированы, причем они не хранятся в секрете и могут быть всем известны.

Чтобы избежать атаки на пароли пользователей, не стоит их сохранять в явном виде на сервере. Пусть $x =$ «password» — некоторый пароль пользователя A с именем «name», причем x представим в виде некоторого числа, например в двоичном представлении в памяти компьютера. Вычислим значение $y = a^x \pmod p$. Тогда в памяти компьютера (на сервере) сохраним пару («name», y). Если пользователь A входит в систему, то он вводит («name», «password»), после чего компьютер вычисляет $\widetilde{y} = a^x \pmod p$, где $x =$ «password». Если значения y и $\widetilde{y}$ совпадают, то пользователь A является законным пользователем. При таком хранении информации о пользователях снижается риск несанкционированного доступа в систему, так как секретный пароль x не хранится в памяти компьютера в явном виде, а попытка вычисления x по значению y заняла бы очень много времени.

Задачу о ПВО можно решить следующим образом. Каждому самолету присвоим имя и секретный пароль x. На станции ПВО будет храниться таблица из имен самолетов и соответствующих секретных паролей, т.е. строки вида («name», x). Бортовой компьютер каждого самолета знает свое имя и секретный пароль. Когда самолет приближается к границе, то сначала он пере-

дает системе ПВО свое имя. Система ПВО находит в таблице паролей значение x, соответствующее данному имени, генерирует случайным образом число a, вычисляет $y = a^x \pmod p$, параллельно передав число a самолету. Бортовой компьютер также вычисляет $\widetilde{y} = a^x \pmod p$ и передает системе ПВО. Если значения y и $\widetilde{y}$ совпадают, то это свой самолет.

9.2. Система Диффи-Хеллмана и ее модификация на эллиптической кривой

Данная криптосхема была придумана в середине 70-х годов американскими учеными Диффи и Хеллманом. Это первая система, которая позволяла защищать информацию без передачи секретных ключей по защищенным каналам. Преимущество данной системы состоит в следующем. Пусть имеется N абонентов, которые хотят организовать между собой секретную переписку. Целесообразно, чтобы каждая пара абонентов имела свой собственный секретный ключ. Тогда всего потребуется $C_N^2 = \dfrac{N(N-1)}{2}$ ключей. Чем больше число N, тем более громоздким и дорогостоящим становится система снабжения абонентов секретными ключами.

Диффи и Хеллман решили эту проблему за счет открытого распространения и вычисления ключей. Опишем данную систему.

Пусть имеются абоненты $A, B, C, \ldots$. Выберем достаточно большое простое число p (например, со значением $k = 1024$ бит) и некоторое число g, которое является первообразным корнем по модулю p (см. определение 1.27 и критерий 1.45), т.е. множество чисел $\{g, g^2, \ldots, g^{p-1}\}$, взятых по модулю числа p, должно представлять собой некоторую перестановку чисел множества $\{1, 2, \ldots, p-1\}$. Числа p и g известны всем абонентам.

Абоненты выбирают достаточно большие числа x_A, x_B, x_C и т.д., которые хранят в секрете. По каждому секретному ключу x вычисляется значение y следующим образом:

$$y_A = g^{x_A} \pmod p,$$

$$y_B = g^{x_B} \pmod p,$$
$$y_C = g^{x_C} \pmod p,$$
$$\dots$$

Значения y_A, y_B, y_C и т.д. размещаются в общедоступном справочнике.

Если два абонента A и B хотят организовать секретную переписку, то они поступают следующим образом. Абонент A находит в справочнике открытый ключ y_B абонента B и вычисляет величину

$$z_{AB} = (y_B)^{x_A} \pmod p.$$

В свою очередь, абонент B вычисляет величину:

$$z_{BA} = (y_A)^{x_B} \pmod p.$$

Заметим при этом, что $z_{AB} = z_{BA}$, так как:

$$z_{AB} = (y_B)^{x_A} = (g^{x_B})^{x_A} = g^{x_B x_A} = g^{x_A x_B} = (y_A)^{x_B} = z_{BA} \pmod p.$$

Таким образом, абоненты A и B получили одно и то же число $z = z_{AB} = z_{BA}$, которое не передавалось по открытой линии связи. При этом если противник знает только значения y_A и y_B, то для нахождения значения z ему требуется решить трудную задачу дискретного логарифмирования.

Абоненты A и B могут использовать число z в качестве секретного ключа для шифрования и расшифрования данных.

Заметим, что дискретные логарифмы сложно вычисляются, когда число $p-1$ содержит один большой простой множитель, например, когда оно представимо в виде $p - 1 = 2q$, где q — простое число. Предположим, что простое число $p = 2q + 1$ выбрано. Следующим шагом требуется выбрать первообразный корень g по модулю p. Для этого можно воспользоваться критерием (теорема 1.45), согласно которому число g будет являться первообразным корнем по модулю p тогда и только тогда, когда:

$$g^2 \not\equiv 1 \pmod p, \quad g^q \not\equiv 1 \pmod p.$$

Модификация системы Диффи-Хэллмана на эллиптических кривых

Безопасность криптосистем на эллиптических кривых ECC (Elliptic Curve Cryptography) как правило, основана на трудности решения задачи дискретного логарифмирования в группе точек эллиптической кривой [59].

Исследования показывают, что в классе криптосистем с открытым ключом криптосистемы на эллиптических кривых превосходят классические криптосистемы на основе модулярной арифметики, как минимум, по двум важным параметрам: степени защищенности в расчете на каждый бит ключа и быстродействию при аппаратной и программной реализации. Наглядно это демонстрирует следующая таблица (длины ключей для ECC и RSA при одинаковой криптостойкости согласно NIST [52]):

ECC key size (Bits)	RSA key size (Bits)	Key ratio	AES key size (Bits)
163	1024	1 : 6	
256	3072	1 : 12	128
384	7680	1 : 20	192
512	15360	1 : 30	256

Любая криптосистема, основанная на дискретном логарифмировании, легко может быть перенесена на эллиптические кривые. В этом случае операция $y = g^x \pmod p$ заменяется на $Y = [x]G$. Пусть q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$ и некоторая точка $G \in E_p(a,b)$ имеет порядок q, т.е. образует циклическую подгруппу порядка q в $(E_p(a,b),+)$:

$$\langle G\rangle = \{G,\ [2]G, \ldots, [q]G = \mathcal{O}\}.$$

Общедоступные параметры системы: p, q, G, $E_p(a,b)$. Абоненты A, B,... выбирают соответствующие секретные ключи x_A, x_B и т.д., не превосходящие числа $q-1$. По каждому секретному ключу вычисляется открытый ключ:

$$Y_A = [x_A]G,$$

$$Y_B = [x_B]G,$$

$$\ldots,$$

которые размещаются в общедоступном справочнике вместе с параметрами системы. Если абоненты A и B хотят организовать секретную связь, то абонент A вычисляет значение:

$$Z_{BA} = [x_A]Y_B,$$

а абонент B вычисляет:

$$Z_{AB} = [x_B]Y_A.$$

При этом $Z_{AB} = Z_{BA} = [x_A x_B]G$. Теперь абоненты A и B могут использовать, например, абсциссу точки Z_{AB} в качестве ключа для секретной переписки.

9.3. Протокол Месси-Омуры и его модификация на эллиптической кривой

Пусть абоненты A, B, C и т.д. хотят организовать между собой секретную переписку. Для этой цели они выбирают достаточно большое простое число p. Каждый абонент независимо друг от друга выбирает себе пару ключей по следующему принципу. Сначала каждый абонент, например A, выбирает произвольным образом натуральное число a_1, которое является взаимно простым с числом $\varphi(p) = p - 1$. После этого абонент A находит значение второго ключа из условия:

$$a_1 x \equiv 1 \pmod{p-1}.$$

Поскольку $(a_1, p-1) = 1$, то существует, и притом единственное, решение данного сравнения (теорема 1.30). Теорема 1.31 дает способ отыскания решения данного сравнения. Пусть a_2 — решение исходного сравнения: $a_1 a_2 \equiv 1 \pmod{p-1}$. Пара чисел a_1, a_2 и является секретным ключом абонента A.

Таким же образом и абонент B сначала выбирает произвольное число b_1, взаимно простое с числом $p-1$, а число b_2 находит из условия $b_1 b_2 \equiv 1 \pmod{p-1}$. После этого у каждого абонента есть своя пара секретных ключей, значения которых знает только их обладатель.

Данные ключи обладают следующим свойством. Пусть m — некоторое целое число, причем $0 < m < p$. Тогда:

$$m^{a_1 a_2} \equiv m \pmod{p}, \quad m^{b_1 b_2} \equiv m \pmod{p}, \ldots$$

Действительно, поскольку $a_1 a_2 \equiv 1 \pmod{p-1}$, то для некоторого целого t будет выполнено равенство $a_1 a_2 = 1 + (p-1)t$. Поэтому:

$$m^{a_1 a_2} = m^{1+(p-1)t} = m \cdot m^{(p-1)t}.$$

Применяя малую теорему Ферма, получаем, что $m^{(p-1)t} \equiv 1 \pmod{p}$, поэтому:

$$m^{a_1 a_2} = m \cdot m^{(p-1)t} \equiv m \pmod{p}.$$

То же самое касается и ключей абонентов B, C и т.д.

Рассмотрим процедуру секретной переписки данных абонентов. Предположим, что абонент A решил отправить сообщение m абоненту B $(0 < m < p)$.

1. Абонент A зашифровывает свое сообщение секретным ключом a_1 : $m_1 \equiv m^{a_1} \pmod{p}$ и сообщение m_1 отправляет абоненту B.

2. Абонент B не может расшифровать сообщение m_1, так как у него нет для этого ключа a_2. Поэтому абонент B зашифровывает сообщение m_1 ключом b_1 : $m_2 \equiv m_1^{b_1} \pmod{p}$ и сообщение m_2 отправляет абоненту A.

3. После этого абонент A зашифровывает сообщение m_2 ключом a_2 : $m_3 \equiv m_2^{a_2} \pmod{p}$ и отправляет сообщение m_3 абоненту B.

4. Теперь абонент B может расшифровать сообщение m_3, т.е. прочитать сообщение m. Для этого ему понадобится ключ b_2 :

   $$m \equiv m_3^{b_2} \pmod{p}.$$

   Данное сравнение следует из того, что $m_3^{b_2} = m^{a_1 b_1 a_2 b_2}$, причем $a_1 b_1 a_2 b_2 = 1 + (p-1)s$ для некоторого s. Поэтому:

   $$m_3^{b_2} = m^{a_1 b_1 a_2 b_2} = m^{1+(p-1)s} = m \cdot m^{(p-1)s} \equiv m \pmod{p}.$$

Пример 9.2. Пусть имеется два абонента A и B, которые решили организовать между собой секретную переписку без передачи ключей. Предположим, что они выбрали для этой цели простое число $p = 29$. Далее абоненты A и B должны выбрать произвольным образом по одному числу, каждое из которых взаимно просто с числом 28. Например, $a_1 = 3$, $b_1 = 5$ для абонентов A и B соответственно.

После этого абонент A решает сравнение $3x \equiv 1 \pmod{28}$. Для этой цели можно воспользоваться теоремой 1.31. Сначала строим такую таблицу:

a_n	9	3
P_n	9	28

Тогда $x = -9 \equiv 19 \pmod{28}$ — решение, поэтому $a_2 = 19$. Аналогично абонент B решает сравнение: $5x \equiv 1 \pmod{28}$. Исходя из таблицы:

a_n	5	1	1	2
P_n	5	6	11	28

получаем, что $x = -11 \equiv 17 \pmod{28}$. Итак, числа 3 и 19 являются секретными ключами абонента A, 5 и 17 — секретными ключами абонента B.

Пусть абонент A хочет передать абоненту B открытый текст $m = 15$. Сначала он шифрует это сообщение своим первым ключом:

$$m_1 = 15^3 \equiv 11 \pmod{29}.$$

Абонент B, получив данное сообщение, не может его расшифровать. Он шифрует это сообщение своим первым ключом 5:

$$m_2 = 11^5 \equiv 14 \pmod{29}.$$

Полученное сообщение абонент A шифрует своим вторым ключом 19:

$$m_3 = 14^{19} \equiv 10 \pmod{29}.$$

Получив сообщение m_3, абонент B, наконец, может расшифровать и получить исходное сообщение m. Делает он это своим

вторым ключом 17:

$$m_4 = 10^{17} \equiv 15 \pmod{29}.$$

Таким образом, $m_4 = m = 15$ — исходное сообщение.

Модификация протокола Месси-Омуры на эллиптической кривой.

Эллиптическая кривая имеет способность приводить константы, на которые умножаются ее точки, по модулю ее порядка (следствие теоремы Лагранжа). Пусть E — эллиптическая кривая порядка n над полем F. Абонент A выбирает первый секретный ключ a_1, взаимно простой с n, и находит такое a_2, что $a_1 a_2 \equiv 1 \pmod{n}$. Абонент B также выбирает такие b_1 и b_2, что $b_1 b_2 \equiv 1 \pmod{n}$.

Абонент A помещает свое сообщение m в некоторую точку эллиптической кривой M достаточно большого порядка. Модифицированный протокол принимает следующий вид.

1. Абонент A вычисляет $M_1 = [a_1]M$ и M_1 отправляет абоненту B.
2. Абонент B вычисляет $M_2 = [b_1]M_1$ и отправляет M_2 абоненту A.
3. Абонент A вычисляет $M_3 = [a_1]M_2$ и M_3 отправляет абоненту B.
4. Абонент B восстанавливает точку M : $M_4 = [b_2]M_3 = M$.

9.4. Вероятностный шифр Эль-Гамаля и его модификация на эллиптической кривой

Пусть абоненты A, B, C и т.д. хотят обмениваться шифрованными сообщениями по открытому каналу связи. Рассмотрим шифр, предложенный Эль-Гамалем, который решает данную задачу, причем, в отличие от шифра Месси-Омуры, данный шифр использует только одну пересылку сообщения.

Для всей группы абонентов выбирается некоторое большое простое число p и первообразный корень g по модулю p, т.е.

такое число g, $1 < g < p$, что множество чисел $\{g, g^2, \ldots, g^{p-1}\}$, взятых по модулю p, является некоторой перестановкой чисел $\{1, 2, \ldots, p-1\}$. Числа p и g передаются всем абонентам.

Далее все абоненты выбирают достаточно большие числа x_A, x_B, x_C и т.д., которые хранят в секрете. С помощью секретного ключа вычисляется значение открытого ключа:

$$y_A = g^{x_A} \pmod{p},\ y_B = g^{x_B} \pmod{p}, \ldots$$

Открытые ключи y_A, y_B и т.д. размещаются в открытом справочнике.

Предположим, что абонент A хочет передать абоненту B сообщение m, где $0 < m < p$.

1. Сначала абонент A генерирует случайным образом некоторое число k, $0 < k < p-1$. Затем он вычисляет числа:

$$c_1 \equiv g^k \pmod{p},$$

$$c_2 \equiv m \cdot y_B^k \pmod{p}.$$

   Пара чисел (c_1, c_2) передается абоненту B.

2. Получив сообщение (c_1, c_2), абонент B вычисляет:

$$\widetilde{m} \equiv c_2 \cdot c_1^{p-1-x_B} \pmod{p}.$$

   При этом $\widetilde{m} = m$, так как:

$$\widetilde{m} = m \cdot y_B^k \cdot g^{k(p-1-x_B)} = m \cdot g^{x_B k} \cdot g^{k(p-1-x_B)} = m \cdot g^{k(p-1)} =$$

$$= m \cdot (g^{p-1})^k \equiv m \cdot 1^k = m \pmod{p}.$$

   В данных выкладках используется теорема Ферма, а именно:

$$g^{p-1} \equiv 1 \pmod{p}.$$

Заметим, что в данной криптосистеме для каждого сообщения m необходимо использовать различные случайные числа k. Действительно, если сообщения m_1 и m_2 шифровались с помощью одного случайного числа k, то из сравнений:

$$c_2 \equiv m_1 \cdot y_B^k \pmod{p},$$

$$\widetilde{c}_2 \equiv m_2 \cdot y_B^k \pmod{p}$$

следует, что:

$$c_2 m_1^{-1} \equiv \widetilde{c}_2 m_2^{-1} \pmod{p}.$$

Следовательно, зная m_1, можно легко вычислить m_2.

Также рекомендуется, чтобы число $p-1$ содержало большой простой делитель.

Вероятностный характер шифрования в системе Эль-Гамаля придает ей большую стойкость, однако при этом объем шифрованного сообщения (c_1, c_2) в два раза превышает объем исходного сообщения m.

Вероятностное шифрование является разновидностью криптосистем с открытым ключом. Данный вид шифрования относят к допускающим неоднозначное вскрытие. При таком шифровании криптограмма сообщения m на ключе k вычисляется с помощью рандомизированного алгоритма: $c = E_k(m, r)$, где r — случайная строка. Это означает, что у каждого сообщения существует, вообще говоря, много криптограмм, вычисленных на одном и том же ключе. Но расшифрование при этом всегда однозначно.

Модификация криптосистемы Эль-Гамаля на эллиптических кривых

Пусть q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$ и некоторая точка $G \in E_p(a,b)$ имеет порядок q.

Общедоступные параметры системы: p, q, G, $E_p(a,b)$. Абонент A выбирает секретный ключ x, $0 < x < q$, и вычисляет открытый ключ $Y = [x]G$.

Абонент B помещает свое сообщение m в некоторую точку эллиптической кривой $M \in E_p(a,b)$ достаточно большого порядка. Модифицированный протокол передачи точки M абоненту A принимает следующий вид.

1. Сначала абонент B генерирует случайным образом некоторое число k, $0 < k < q$. Затем он вычисляет точки эллип-

тической кривой:

$$C_1 = [k]G,$$
$$C_2 = M + [k]Y.$$

Пара точек (C_1, C_2) передается абоненту A.

2. Получив сообщение (C_1, C_2), абонент A вычисляет:

$$\widetilde{M} = C_2 + [q - x]C_1.$$

При этом $\widetilde{M} = M$, так как:

$$C_2 + [q - x]C_1 = M + [kx]G + [k]([q]G) - [kx]G = M.$$

Можно также преобразовать предыдущий протокол без предварительного преобразования сообщения m в некоторую точку эллиптической кривой $M \in E_p(a, b)$. Пусть $0 < m < q$.

1. Сначала абонент B генерирует случайным образом некоторое число k, $0 < k < q$. Затем он вычисляет точки эллиптической кривой:

$$C_1 = [k]G,$$
$$C_2 = [k]Y = (x_2, y_2)$$

и значение $c = m + x_2 \pmod q$. Пара $(C_1,)$ передается абоненту A.

2. Получив сообщение $(C_1,)$, абонент A вычисляет:

$$C_2 = [x]C_1 = (x_2, y_2), \quad m = c - x_2 \pmod q.$$

9.5. Шифр RSA

Данный шифр назван в честь его разработчиков Риверса, Шамира и Адлемана и является одним из самых широко используемых.

Как было показано ранее, шифры Месси-Омуры и Эль-Гамаля решают задачу обмена шифрованными сообщениями по открытому каналу, но в первом случае сообщение передается три раза от одного абонента к другому, а во втором случае объем шифртекста в два раза превышает объем исходного сообщения. Система RSA лишена подобных недостатков.

Данная система базируется на следующих фактах из теории чисел:

1. Сравнительно легко проверить число на простоту.
2. Очень трудно разложить достаточно большое число на множители (задача факторизации).

Пусть имеются абоненты A, B, C и т.д. Каждый абонент выбирает два больших простых числа p и q и вычисляет:

$$n = p \cdot q,$$

$$\varphi(n) = (p-1) \cdot (q-1),$$

где $\varphi(n)$ — функция Эйлера. Затем выбирается некоторое число e, $e < \varphi(n)$, взаимно простое с $\varphi(n)$, и находится решение сравнения:

$$ed \equiv 1 \pmod{\varphi(n)},$$

которое по теореме 1.30 имеет единственное решение.

Тогда секретными ключами абонентов A, B, C и т.д. будут числа d_A, d_B, d_C,..., а открытыми ключами будут пары чисел (e_A, n_A), (e_B, n_B), (e_C, n_C),...

Пусть m — некоторое сообщение, которое абонент A хочет отправить абоненту B, причем $m < n_B$. Тогда абонент A находит в справочнике открытый ключ абонента B и вычисляет:

$$m_1 \equiv m^{e_B} \pmod{n_B}.$$

После чего шифрованное сообщение m_1 по открытому каналу передается абоненту B, который, получив m_1, вычисляет:

$$m_2 \equiv m_1^{d_B} \pmod{n_B}.$$

При этом $m = m_2$. Действительно, так как:

$$e_B d_B \equiv 1 \pmod{\varphi(n_B)},$$

то найдется такое число t, что:

$$e_B d_B = 1 + t\varphi(n_B).$$

Поэтому:

$$m_2 = m^{e_B d_B} = m^{1+t\varphi(n_B)} = m \cdot (m^{\varphi(n_B)})^t \equiv m \cdot 1^t = m \pmod{n_B}.$$

В данном равенстве использовалась теорема Эйлера (теорема 1.27), а именно:

$$m^{\varphi(n_B)} \equiv 1 \pmod{n_B}, \quad (m, n_B) = 1.$$

Замечание 9.2. Может оказаться так, что числа m и $n = n_B$ не взаимно просты. В этом случае теорему Эйлера применить нельзя. Только это событие маловероятно при больших p и q, так как оно не больше числа:

$$1 - \frac{\varphi(n)}{n} = 1 - \frac{(p-1)(q-1)}{pq} = \frac{1}{p} + \frac{1}{q} - \frac{1}{pq}.$$

Заметим, что числа p и q должны быть выбраны таким образом, чтобы разложение числа n было очень сложным в вычислительном плане. Для этого данные числа должны удовлетворять следующим условиям:

1) числа p и q должны быть достаточно большими;

2) числа p и q должны быть такими, что число $(p-1, q-1)$ является небольшим, например, $(p-1, q-1) = 2$;

3) числа p и q должны быть сильно простыми числами (*сильно простым* называется такое число r, что число $r-1$ имеет большой простой делитель s и число $r+1$ имеет большой простой делитель t, причем число $s-1$ обладает большим простым делителем и число $t-1$ обладает большим простым делителем).

Заметим также важность того, чтобы каждый абонент выбирал собственную пару простых чисел p и q. В противном же случае при совпадении, например, чисел n_A и n_B абонент A может прочитать зашифрованное сообщение, переданное абонентом C абоненту B. То же самое может сделать и абонент B. При этом никаких ограничений на число e не накладывается, оно может быть одинаковым у всех абонентов. Часто рекомендуется брать $e = 3$, только при этом ни одно из чисел $p-1$ и $q-1$ не должно делиться на 3. Тогда шифрование выполняется максимально быстро.

9.6. Рюкзачная криптосистема Меркла-Хеллмана

«Проблема рюкзака» может быть сформулирована следующим образом. Пусть имеется некоторое множество натуральных чисел $A = \{a_1, a_2, \ldots, a_n\}$ и некоторое натуральное число S. Требуется проверить, имеется ли такое подмножество в A, сумма элементов которого была бы равна S. Также можно сформулировать эквивалентную задачу: существует ли такой двоичный набор x_1, x_2,..., x_n, состоящий из нулей и единиц, что:

$$x_1 a_1 + x_2 a_2 + \ldots + x_n a_n = S.$$

Пусть M — некоторое сообщение в некотором алфавите. Если каждый символ данного алфавита закодировать двоичным кодом фиксированной длины, то сообщение M можно представить в виде двоичной последовательности $M = M_1 M_2 \ldots$, где все $M_i \in \{0, 1\}$. После этого двоичная последовательность $M_1 M_2 \ldots$ разбивается на блоки длиной n и шифруется в последовательность $S_1 S_2 \ldots$ таким образом:

$$S_1 = \sum_{i=1}^{n} M_i a_i, \quad S_2 = \sum_{i=1}^{n} M_{n+i} a_i \quad \text{и т.д.}$$

Понятно, что процесс шифрования происходит очень легко, а вот расшифрование предполагает большие затраты времени, например, можно попытаться расшифровать полным перебором (2^n вариантов). До сих пор не известно, имеет ли решение «проблемы рюкзака» с полиномиальной сложностью.

Построение систем шифрования на основе проблемы рюкзака заключается в следующем. Сначала выделяется некоторый подкласс задач об укладке рюкзака, которые решаются сравнительно легко. Затем задачи этого класса «маскируются» путем преобразования некоторых параметров под общий случай.

Рассмотрим пример легко решаемой задачи, которую в 1978 г. предложили Р. Меркль и М. Хеллман.

Последовательность натуральных чисел $b_1, b_2, \ldots, b_n$ назовем *супервозрастающей*, если для любого $i = 2, 3, \ldots, n$ выполняет-

ся неравенство

$$b_i > b_1 + b_2 + \ldots + b_{i-1}.$$

В этом случае очевидным является следующее предложение.

Предложение 9.2. Пусть $b_1, b_2, \ldots, b_n$ — некоторая супервозрастающая последовательность натуральных чисел и пусть $S = x_1 b_1 + \ldots + x_n b_n$, где x_1, x_2,..., x_n — некоторый двоичный набор. Тогда элемент b_n входит в сумму S (т.е. $x_n = 1$) тогда и только тогда, когда $b_n \leqslant S$.

На основе данного предложения проблема рюкзака для супервозрастающей последовательности может быть решена с помощью следующей процедуры.

1. Полагаем $i = n$.
2. Если $i \geqslant 1$, то сравниваем b_i и S. Если $b_i \leqslant S$, то $x_i := 1$, $S := S - b_i$, в противном случае $x_i := 0$.
3. $i := i - 1$ и переходим к шагу 2.

Преобразуем супервозрастающую последовательность b_1, b_2,..., b_n в последовательность $a_1, a_2, \ldots, a_n$ следующим образом. Выберем модуль m таким образом, чтобы $m > \sum_{i=1}^{n} b_i$. Зафиксируем произвольным образом такое число t, $1 < t < m - 1$, что $(t, m) = 1$. Каждый элемент последовательности $a_1, a_2, \ldots, a_n$ вычислим по правилу $a_i \equiv b_i \cdot t \pmod{m}$.

Открытым ключом будет являться набор $a_1, a_2, \ldots, a_n$, секретным ключом — набор $m, t, b_1, b_2, \ldots, b_n$.

Пусть $M = M_1 M_2 \ldots$ — некоторое сообщение, где $M_i \in \{0, 1\}$. Разобьем двоичную последовательность $M_1 M_2 \ldots$ на блоки длиной n и зашифруем с помощью открытого ключа $a_1, a_2, \ldots, a_n$ в последовательность $S_1 S_2 \ldots$:

$$S_1 = M_1 a_1 + \ldots + M_n a_n, \quad S_2 = M_{n+1} a_1 + \ldots + M_{2n} a_n \quad \text{и т.д..}$$

Абонент, обладающий секретным ключом $m, t, b_1, b_2, \ldots, b_n$, получив сообщение $S_1 S_2 \ldots$, расшифровывает его следующим

образом. Сначала данный абонент преобразовывает шифрованную последовательность $S_1S_2\ldots$ в последовательность $\widetilde{S}_1\widetilde{S}_2\ldots$, где $\widetilde{S}_i \equiv uS_i \pmod m$, u — решение сравнения $tu \equiv 1 \pmod m$:

$$\widetilde{S} \equiv uS = u(M_1a_1 + \ldots + M_na_n) \equiv u(M_1b_1t + \ldots + M_nb_nt) \equiv$$

$$\equiv ut(M_1b_1 + \ldots + M_nb_n) \equiv M_1b_1 + \ldots + M_nb_n \pmod m.$$

Затем для каждого значения $\widetilde{S}_i$ решается проблема рюкзака для супервозрастающей последовательности $b_1, b_2, \ldots, b_n$. Этим самым полностью восстанавливается исходное сообщение $M = M_1M_2\ldots$

Заметим, что данная криптосистема не является стойкой, так как доказано, что существует алгоритм полиномиальной сложности, который может быть использован противником для получения открытого текста по шифртексту.

9.7. Рюкзачная криптосистема Шора-Ривеста на основе конечных полей

Задача о рюкзаке, лежащая в основе базисного варианта систем с открытым распределением ключей (рассматривалась в главе 9), имеет низкую плотность, в том смысле, что компоненты рюкзачного вектора располагается на отрезке от 1 до n очень редко. Плотность рюкзака определяется следующим образом:

$$\rho(A) = \frac{n}{\log_2(\max\ A)}.$$

Например, для сверхрастущего вектора $A = (1, 2, 4, 8, 16, 32, 64)$ $\rho(A) = 7/6$.

Увеличить плотность рюкзачного вектора можно с использованием полей Галуа $GF(p^d)$. Напомним, что с точностью до изоморфизма поле $\mathbb{Z}_p$ является подполем в $GF(p^d)$.

Пусть g — образующий элемент мультипликативной группы поля $GF(p^d)$ (теорема 2.18, причем таких образующих ровно $\varphi(p^d-1)$). Образующий элемент можно рассматривать как основание дискретного логарифма log от элементов поля. Рассмотрим пример. Пусть $p = 3$, $d = 2$. Поле $GF(p^d)$ состоит из эле-

ментовЖ

$$0,\ 1,\ 2,\ \alpha,\ \alpha+1,\ \alpha+2,\ 2\alpha,\ 2\alpha+1,\ 2\alpha+2,$$

где α соответствует неприводимому многочлену x^2-x-1 над полем $\mathbb{Z}_3$. В этом случае $\alpha^2=\alpha+1$. В поле $GF(3^2)$ элемент α является образующим, так как:

$$\begin{aligned}
&\alpha^2=\alpha+1,\\
&\alpha^3=\alpha\alpha^2=\alpha(\alpha+1)=2\alpha+1,\\
&\alpha^4=\alpha\alpha^3=\alpha(2\alpha+1)=2,\\
&\alpha^5=\alpha\alpha^4=2\alpha,\\
&\alpha^6=\alpha\alpha^5=2\alpha^2=2\alpha+2,\\
&\alpha^7=\alpha\alpha^6=\alpha(2\alpha+2)=\alpha+2,\\
&\alpha^8=\alpha\alpha^7=\alpha^2+2\alpha=1.
\end{aligned}$$

Запишем полученные степени элемента α в виде таблицы:

i	1	2	3	4	5	6	7	8
α^i	α	$\alpha+1$	$2\alpha+1$	2	2α	$2\alpha+2$	$\alpha+2$	1

Из приведенной таблицы видно, что α является образующим элементом поля $GF(3^2)$. Заметим, что из теоремы 2.18 и предложения 2.15 следует, что в поле $GF(3^2)$ ровно $\varphi(8)=4$ образующих элементов, причем это следующие элементы: α, $\alpha^3=2\alpha+1$, $\alpha^5=2\alpha$, $\alpha^7=\alpha+2$.

Данная таблица может быть представлена как таблица дискретных логарифмов. Для этого в верхней строке запишем ненулевые элементы поля в упорядоченном виде, а в нижней значения степеней образующего элемента, при котором получаем данный элемент поля:

y	1	2	α	$\alpha+1$	$\alpha+2$	2α	$2\alpha+1$	$2\alpha+2$
$\log_\alpha y$	8	4	1	2	7	5	3	6

Вычисление дискретных логарифмов считается трудной задачей, как и задача факторизации. Таблица логарифмов может использоваться для выполнения умножения и деления элементов поля. Например:

$$\log_\alpha\big((\alpha+1)(2\alpha+1)\big)=\log_\alpha(\alpha+1)+\log_\alpha(2\alpha+1)=2+3=5,$$

поэтому $(\alpha+1)(2\alpha+1)=2\alpha$.

$\log_\alpha\big((\alpha+2)2\alpha\big)=\log_\alpha(\alpha+2)+\log_\alpha(2\alpha)=7+5=12\equiv 4 \pmod 8$,

поэтому $(\alpha+2)2\alpha=2$. Все операции выполняются по модулю p^d-1, в данном примере по модулю 8. Также:

$$\log_\alpha\big((\alpha+1)/(\alpha+2)\big)=\log_\alpha(\alpha+1)-\log_\alpha(\alpha+2)=$$
$$=2-7=-5\equiv 3 \pmod 8,$$

поэтому $(\alpha+1)/(\alpha+2)=2\alpha+1$.

Рассмотрим вспомогательную задачу. Пусть имеется рюкзачный вектор $A=(a_1,a_2,\ldots,a_n)$. Рассмотрим различные суммы из d компонент. Задача состоит в том, чтобы для заданных n и d построить такой вектор A, чтобы все суммы из d элементов были бы попарно различны. Для рюкзаков низкой плотности это сделать легко: $a_i=d^{i-1}$, $i=1,\ldots,n$ (теорема 1.2). Например:

$$A=(1,2,4,8,16,32,64),\quad n=7,\ d=2.$$

Такое построение векторов A соответствует рюкзакам низкой плотности, так как они являются сверхрастущими.

Нам понадобится следующее утверждение.

Теорема 9.1 (**Bose-Chowla**)**.** Для любого простого p и любого натурального $d\geqslant 2$ всегда можно построить рюкзачный вектор $A=(a_1,a_2,\ldots,a_p)$, удовлетворяющий следующим условиям:

1) $1\leqslant a_i\leqslant p^d-1$, $i=1,\ldots,p$;

2) если x_i и y_i — неотрицательные целые числа, $i=1,\ldots,p$, такие, что:

$$(x_1,x_2,\ldots,x_p)\neq(y_1,y_2,\ldots,y_p) \text{ как упорядоченные наборы :} \tag{9.2}$$
$$\sum_{i=1}^p x_i\leqslant d-1,\ \sum_{i=1}^p y_i\leqslant d-1 \text{ или } \sum_{i=1}^p x_i=\sum_{i=1}^p y_i=d,$$

то:

$$\sum_{i=1}^p x_ia_i\not\equiv\sum_{i=1}^p y_ia_i \pmod{p^d-1}. \tag{9.3}$$

Доказательство. Рассмотрим расширение $GF(p) \subset GF(p^d)$. Пусть $\alpha \in GF(p^d)$ — корень неприводимого многочлен степени d над полем $GF(p)$, g — образующий элемент поля $GF(p^d)$. Рассмотрим следующие элементы:

$$a_i = \log_g(\alpha + i - 1), \quad i = 1, \ldots, p. \tag{9.4}$$

Понятно что для данных чисел будет выполнен пункт 1. Покажем справедливость второго пункта. Предположим, что для векторов $(x_1, x_2, \ldots, x_p)$, $(y_1, y_2, \ldots, y_p)$ выполнено условие (9.2), но при этом:

$$\sum_{i=1}^{p} x_i a_i \equiv \sum_{i=1}^{p} y_i a_i \pmod{p^d - 1}.$$

$$g^{\sum_{i=1}^{p} x_i a_i} = g^{\sum_{i=1}^{p} y_i a_i}.$$

Так как:

$$\sum_{i=1}^{p} x_i a_i = \log_g\bigl(\alpha^{x_1}(\alpha+1)^{x_2}\ldots(\alpha+p-1)^{x_p}\bigr),$$

$$\sum_{i=1}^{p} y_i a_i = \log_g\bigl(\alpha^{y_1}(\alpha+1)^{y_2}\ldots(\alpha+p-1)^{y_p}\bigr),$$

то:

$$\alpha^{x_1}(\alpha+1)^{x_2}\ldots(\alpha+p-1)^{x_p} = \alpha^{y_1}(\alpha+1)^{y_2}\ldots(\alpha+p-1)^{y_p}.$$

Так как $x_i \neq y_i$ для некоторого i, то обе части последнего равенства представляют собой разные многочлены от переменной α. Разность этих двух многочленов представляет собой ненулевой многочлен над $GF(p)$ степени, меньше чем d, корнем которого является α. Это противоречит тому, что α — алгебраический элемент степени d над $GF(p)$. □

Замечание 9.3. В формулировке и доказательстве теоремы простое число p можно заменить на степень простого числа: p^s для любого $s \in \mathbb{N}$.

Начнем с более «облегченного» описания криптосистемы Шора-Ривеста, постепенно усложняя ее.

Криптосистема. Опишем рюкзачную систему [50], которая до сих пор пользуется доверием и не поддается вскрытию.

Пусть p — простое (или степень простого числа), d — натуральное число, α — алгебраический элемент поля $GF(p^d)$ над $GF(p)$ степени d, g — образующий элемент поля $GF(p^d)$. Напомним, что если $p^d - 1 = p_1^{k_1} \dots p_n^{k_n}$ — каноническое разложение числа $p^d - 1$, то ненулевой элемент $g \in GF(p^d)$ является образующим элементом тогда и только тогда, когда:

$$g^{\frac{p^d-1}{p_i}} \neq 1, \quad i = 1, \dots, n.$$

Предварительно вычисляется таблица степеней:

i	1	2	$\dots$	p^d-1
g^i	g	g^2	$\dots$	g^{p^d-1}

По формуле (9.4) вычисляем $a_i = \log_g(\alpha + i - 1)$, $i = 1, \dots, p$. Открытым ключом в данной криптосистеме являются $A = (a_1, \dots, a_p)$, p, d, секретный ключ — α, неприводимый многочлен $f(x)$ степени d над $GF(p)$ и g.

При построении криптосистемы исходный текст должен состоять из p-разрядных блоков, в каждом из которых сумма всех компонент равна d. Для того, чтобы обеспечить это условие, необходимо после кодирования символов сообщений произвести перекодирование равновесными кодами длины p веса d.

Пусть $(x_1, \dots, x_p)$ — очередной символ открытого текста веса d. Данный символ шифруется по правилу

$$s = \sum_{i=1}^{p} a_i x_i \pmod{p^d - 1},$$

где s — очередной символ шифрованного текста.

Рассмотрим как расшифровать s владельцем секретного ключа, т.е. найти вектор $(x_1, \dots, x_p)$. Так как:

$$s = \sum_{i=1}^{p} a_i x_i = \log_g\bigl(\alpha^{x_1}(\alpha+1)^{x_2} \dots (\alpha+p-1)^{x_p}\bigr),$$

то:

$$g^s = \alpha^{x_1}(\alpha+1)^{x_2} \dots (\alpha+p-1)^{x_p}.$$

Разложив на линейные множители (по переменной α) многочлен:

$$f(\alpha) + g^s = (\alpha + \beta_1)(\alpha + \beta_2)\dots(\alpha + \beta_d),\ \ \beta_i \in GF(p),$$

получаем значения $x_1, \dots, x_p$.

Пример 9.3. Выберем образующий элемент $g = 2\alpha + 1$ поля $GF(3^2)$. Учитывая, что $\alpha^3 = 2\alpha + 1$, выпишем степени элемента $2\alpha + 1$ в виде таблицы:

i	1	2	3	4	5	6	7	8
$(2\alpha+1)^i$	$2\alpha+1$	$2\alpha+2$	α	2	$\alpha+2$	$\alpha+1$	2α	1

Учитывая формулу (9.4):

$$\begin{aligned} a_1 &= \log_{2\alpha+1}(\alpha) = 3, \\ a_2 &= \log_{2\alpha+1}(\alpha+1) = 6, \\ a_3 &= \log_{2\alpha+1}(\alpha+2) = 5. \end{aligned}$$

Поэтому $A = (3, 6, 5)$. Каждый символ шифруемого сообщения нужно перекодировать векторами длины 3 веса 2:

$$(0,0,2),\ (0,1,1),\ (0,2,0),\ (1,0,1),\ (1,1,0),\ (2,0,0). \qquad (9.5)$$

Открытым ключом в данной криптосистеме являются A, p, d, секретный ключ — α, неприводимый многочлен $x^2 - x - 1$ и g.

После шифрования символов (9.5) преобразованного открытого текста получаем:

$$(3\ 6\ 5) \cdot \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} = (2,3,4,0,1,6).$$

В данном случае производились вычисления по модулю $p^d - 1$ (в нашем случае по модулю 8). Таким образом, при шифровании получаем такое соответствие:

$$(0,0,2) \to 2,\ \ (0,1,1) \to 3,\ \ (0,2,0) \to 4,$$

$$(1,0,1) \to 0,\ \ (1,1,0) \to 1,\ \ (2,0,0) \to 6.$$

Рассмотрим процесс расшифрования шифртекста владельцем секретного ключа:

$$2 = x_1a_1 + x_2a_2 + x_3a_3 =$$

$$= x_1 \log_{2\alpha+1}(\alpha) + x_2 \log_{2\alpha+1}(\alpha+1) + x_3 \log_{2\alpha+1}(\alpha+2) =$$

$$= \log_{2\alpha+1}\bigl(\alpha^{x_1}(\alpha+1)^{x_2}(\alpha+2)^{x_3}\bigr).$$

Цель — найти вектор (x_1, x_2, x_3). Из последних равенств следует, что:

$$\alpha^{x_1}(\alpha+1)^{x_2}(\alpha+2)^{x_3} = (2\alpha+1)^2 = 2\alpha+2.$$

Причем к полученным значениям можно прибавлять $\alpha^2-\alpha-1$. Поэтому:

$$2\alpha+2 = \alpha^2+\alpha+1 = \alpha^2-2\alpha+1 = (\alpha-1)^2 = (\alpha+2)^2,$$

что означает $x_1 = 0$, $x_2 = 0$, $x_3 = 2$, $2 \to (0,0,2)$. Аналогично:

$$(2\alpha+1)^3 = \alpha = \alpha^2-1 = (\alpha+1)(\alpha+2),\ 3 \to (0,1,1),$$

$$(2\alpha+1)^4 = 2 = \alpha^2-\alpha+1 = (\alpha+1)^2,\ 4 \to (0,2,0),$$

$$(2\alpha+1)^0 = 1 = \alpha^2-\alpha = \alpha(\alpha+2),\ 0 \to (1,0,1),$$

$$(2\alpha+1)^1 = 2\alpha+1 = \alpha^2+\alpha = \alpha(\alpha+1),\ 1 \to (1,1,0),$$

$$(2\alpha+1)^6 = \alpha+1 = \alpha^2,\ 6 \to (2,0,0).$$

Для усложнения криптоанализа в данной системе вводится перемешивание σ и сдвиг (шум) h. Эти значения являются дополнительной лазейкой для легального пользователя.

В данном случае после составления рюкзачного вектора A генерируется сдвиг $0 \leqslant h \leqslant p^d-1$ и все его компоненты преобразуются по правилу:

$$b_i = a_i + h \pmod{p^d-1}, \quad i = 1, \ldots, p.$$

Открытым ключом будут являться $B = (b_1, \ldots, b_p)$, p, d. Пусть $(x_1, \ldots, x_p)$ — очередной символ открытого текста веса d. Он шифруется по правилу:

$$s = \sum_{i=1}^{p} x_i b_i \pmod{p^d-1}.$$

Рассмотрим процесс расшифрования s легальным пользователем. Имеем:

$$s = \sum_{i=1}^{p} b_i x_i = \sum_{i=1}^{p} a_i x_i + \sum_{i=1}^{p} h x_i =$$

$$= \log_g \left(\alpha^{x_1} (\alpha + 1)^{x_2} \dots (\alpha + p - 1)^{x_p} \right) + hd.$$

Поэтому:

$$g^{s-hd} = \alpha^{x_1} (\alpha + 1)^{x_2} \dots (\alpha + p - 1)^{x_p}.$$

Далее вектор $(x_1, \dots, x_p)$ восстанавливается по описанному выше алгоритму.

Рассмотрим самый общий случай шифрования с использованием полей Галуа. Пусть $\sigma \in S_p$, где S_p — симметрическая группа степени p, h — целое число. После составления рюкзачного вектора A преобразуем его следующим образом: $B = (a_{\sigma(1)}, \dots, a_{\sigma(p)})$, после этого введем шумовую составляющую в виде сдвига h :

$$c_i = b_i + h \pmod{p^d - 1}, \quad i = 1, \dots, p.$$

Открытым ключом будут являться $C = (c_1, \dots, c_p)$, p, d.

Глава 10. Криптографические хеш-функции

10.1. Сбалансированные функции

Определение 10.1. Пусть X, Y — некоторые конечные множества. Отображение $h : X \to Y$ называется *сбалансированным*, если для любых $y_1, y_2 \in Y$ мощности полных прообразов элементов y_1 и y_2 совпадают: $|h^{-1}(y_1)| = |h^{-1}(y_2)|$.

Для любого отображения $f : X \to Y$ выполнены такие равенства:

$$X = \bigcup_{y \in Y} f^{-1}(y), \quad |X| = \sum_{y \in Y} |f^{-1}(y)|.$$

Поэтому если h является сбалансированным отображением, то для любого фиксированного $y \in Y$ верно такое равенство:

$$|h^{-1}(y)| = \frac{|X|}{|Y|}.$$

Если $|X| = |Y|$, то сбалансированность отображения h эквивалентна его биективности.

Предложение 10.1. Для произвольной функции $h : X \to Y$, $|X| < +\infty$, $|Y| < +\infty$, верны следующие условия.

1. Случайным равновероятным образом выберем элемент декартова произведения $(x, y) \in X \times Y$. Тогда вероятность того, что $h(x) = y$ равна $1/|Y|$.

2. Предположим, что h — сбалансированное отображение. Зафиксируем произвольный элемент $y_0 \in Y$. Случайным равновероятным образом выберем $x \in X$. Тогда вероятность того, что $h(x) = y_0$ равна $1/|Y|$.

3. Зафиксируем произвольный элемент $x_0 \in X$. Случайным равновероятным образом выберем $y \in Y$. Тогда вероятность того, что $h(x_0) = y$ равна $1/|Y|$.

Доказательство. Пространством элементарных исходов в нашем эксперименте будет множество $\Omega = X \times Y$.

Пусть A — матрица размером $|X| \times |Y|$ над множеством $\{0, 1\}$, в которой строки занумерованы элементами множества X, а столбцы — элементами множества Y, причем элемент матрицы A с номером (x, y), $(x, y) \in X \times Y$, равен 1, если $h(x) = y$, иначе 0. Понятно, что в каждой строке матрицы A имеется ровно одна единица. Если h — сбалансированное отображение, то в каждом столбце — ровно $\dfrac{|X|}{|Y|}$ единиц. Поэтому вероятность того, что $h(x) = y$ равна отношению количества всех единиц матрицы A к общему количеству ее элементов:

$$P(h(x) = y) = \frac{|X|}{|X| \cdot |Y|} = \frac{1}{|Y|}.$$

Пусть $y_0 \in Y$. Так как в каждом столбце матрицы A для сбалансированного отображения ровно $|X|/|Y|$ единиц, то $P(h(x) = y_0) = 1/|Y|$.

Пусть $x_0 \in Y$. Так как в каждой строке матрицы A имеется ровно одна единица, то $P(h(x_0) = y) = 1/|Y|$. □

Предложение 10.2. Пусть $h : X \to Y$ — некоторая сбалансированная функция. Случайным равновероятным образом выберем два элемента $x_1, x_2 \in X$, $x_1 \neq x_2$. Тогда:

1) если $|X| > |Y|$, то вероятность того, что $h(x_1) = h(x_2)$ не превосходит числа $\dfrac{1}{|Y|}$;

2) если $|X| = |Y|$, то $P(h(x_1) = h(x_2)) = 0$.

Доказательство. 1. Пространством элементарных исходов в данном эксперименте будет являться множество всех сочетаний множества X из $|X|$ по 2 элемента:

$$\Omega = \{(x_1, x_2) \mid x_1, x_2 \in X,\ x_1 < x_2\}.$$

При этом $|\Omega| = C^2_{|X|}$.

Зафиксируем некоторый элемент $y \in Y$. Вероятность того, что для случайно выбранной пары $(x_1, x_2) \in \Omega$ выполнено равенство $h(x_1) = h(x_2) = y$, вычисляется по такой формуле:

$$P(h(x_1) = h(x_2) = y) = \frac{C^2_{|h^{-1}(y)|}}{C^2_{|X|}}.$$

Так как отображение h сбалансировано, то $|h^{-1}(y)| = \dfrac{|X|}{|Y|}$. Следовательно:

$$P(h(x_1) = h(x_2) = y) = \frac{C^2_{\frac{|X|}{|Y|}}}{C^2_{|X|}} = \frac{\frac{|X|}{|Y|} \cdot (\frac{|X|}{|Y|} - 1)}{|X| \cdot (|X| - 1)} =$$

$$= \frac{|X| - |Y|}{|Y|^2 \cdot (|X| - 1)}.$$

Исходя из данного равенства, получаем:

$$P(h(x_1) = h(x_2)) = \sum_{y \in Y} P\left(h(x_1) = h(x_2) = y\right) =$$

$$= |Y| \cdot \frac{|X| - |Y|}{|Y|^2 \cdot (|X| - 1)} = \frac{|X| - |Y|}{|Y| \cdot (|X| - 1)} \leqslant$$

$$\leqslant \frac{|X| - |Y|}{|Y| \cdot (|X| - |Y|)} = \frac{1}{|Y|}.$$

2. Так как h — сбалансированное отображение и $|X| = |Y|$, то h является биекцией. Следовательно, $P(h(x_1) = h(x_2)) = 0$ при $x_1 \neq x_2$. □

Предложение 10.3. Пусть $h : X \to Y$ — некоторая сбалансированная функция, $x_0 \in X$. Случайным равновероятным образом выберем элемент $x \in X$, $x \neq x_0$. Тогда:

1) если $|X| > |Y|$, то:

$$P(h(x) = h(x_0)) \leqslant \frac{1}{|Y|};$$

2) если $|X| = |Y|$, то $P(h(x) = h(x_0)) = 0$.

Доказательство. В данном случае пространством элементарных исходов будет являться такое множество: $\Omega = X \backslash \{x_0\}$.

1. Обозначим через K_x множество всех таких $x \in X$, для которых выполнено равенство $h(x) = h(x_0)$. В силу сбалансированности отображения h имеем $|K_x| = \dfrac{|X|}{|Y|}$. Поэтому:

$$P(h(x) = h(x_0)) = \frac{|K_x| - 1}{|X| - 1} = \frac{\frac{|X|}{|Y|} - 1}{|X| - 1} = \frac{|X| - |Y|}{|Y| \cdot (|X| - 1)} \leqslant$$

$$\leqslant \frac{|X| - |Y|}{|Y| \cdot (|X| - |Y|)} = \frac{1}{|Y|}.$$

2. Следует из биективности отображения h. □

10.2. Хеш-функции и целостность данных

Пусть A — некоторый конечный алфавит, в котором записываются сообщения (как правило, $A = \{0, 1\}$), A^* — множество всех конечных слов (сообщений) в алфавите A, $X \subseteq A^*$, $Y = V_m$ — множество двоичных векторов длины m.

Определение 10.2. *Хеш-функцией* называют всякую легко вычисляемую функцию $h : X \to Y$. Значение хеш-функции называется ее *сверткой*.

Замечание 10.1. Основным требованием, предъявляемым к хеш-функциям, является *равномерность* распределения их значений при случайном выборе значений аргументов. Для конечного множества X при случайном и равновероятном выборе сообщений это условие эквивалентно сбалансированности отображения h.

Пару сообщений с одинаковыми значениями хеш-функции называют *коллизией*.

Пример 10.1. 1. Пусть $A = V_n$ и $x_1 x_2 \dots x_l \in X$. Определим функцию h следующим образом:

$$h(x_1 x_2 \dots x_l) = x_1 + x_2 + \dots + x_l \pmod{2^m}.$$

Данная хеш-функция является контрольной суммой для сообщения и может использоваться для обнаружения ошибок в сообщении.

2. Пусть $A = V_n$, $Y = V_m$, $X = V_{mn}$. Рассмотрим две матрицы:

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Пусть C_1, C_2, C_3,..., C_{2^n-1} — произвольный набор матриц из множества $\{B_1, B_2\}$. Матрицы C_1, ..., C_{2^n-1} выпишем последовательно одна под другой. В результате этого получится матрица B порядка $2^n \times 2$, в которой строки занумерованы элементами множества V_n, а столбцы — элементами множества V_2.

На основе матрицы B построим сбалансированное отображение $h : X \to Y$ следующим образом. Пусть $x = x_1x_2 \dots x_m \in X$, где все $x_i \in V_n$. Положим $h(x_i) = 0$, $x_i \in V_n$, если на пересечении строки с номером x_i и столбца с номером 0 матрицы B находится единица, в противном случае, $h(x_i) = 1$. После этого полагаем $h(x_1x_2 \dots x_m) = h(x_1)h(x_2) \dots h(x_m)$.

Пусть $h : X \to Y$ — некоторая хеш-функция (со свойством равномерности распределения значений), $|X| < +\infty$, $x \in X$. В качестве x может быть некоторое сообщение, некоторые данные на компьютере и т.д. Для того чтобы контролировать целостность данных x, поступают следующим образом. Вычисляется значение $y = h(x)$ и данные хранятся или передаются в виде (x, y).

Предположим, что данные (x, y) были каким-либо случайным образом изменены. Тогда они могут принять такой вид: 1. $(\widetilde{x}, y)$, 2. $(x, \widetilde{y})$, 3. $(\widetilde{x}, \widetilde{y})$, где $\widetilde{x} \in X$, $\widetilde{y} \in Y$, $x \neq \widetilde{x}$, $y \neq \widetilde{y}$.

Вычислим вероятность того, что данные были изменены, но этот факт остался незамеченным.

1. Рассмотрим вариант $(\widetilde{x}, y)$. Вычислим вероятность того, что $h(\widetilde{x}) = y$, т.е. с какой вероятностью факт изменения данных в сообщении x не будет обнаружен.

Так как в данном случае значение y фиксировано, а изменилось только $x \in X$, то пространством элементарных исходов

будет являться множество $X\backslash\{x\}$. Событию $h(\widetilde{x}) = y$ благоприятствуют такие элементарные исходы $\widetilde{x} \in X\backslash\{x\}$, при которых $h(\widetilde{x}) = y$. Всего таких исходов $|h^{-1}(y)| - 1$. В силу сбалансированности отображения h имеем такое равенство:

$$|h^{-1}(y)| - 1 = \frac{|X|}{|Y|} - 1.$$

Следовательно:

$$P(h(\widetilde{x}) = y) = \frac{|h^{-1}(y)| - 1}{|X| - 1} = \frac{|X| - |Y|}{|Y| \cdot (|X| - 1)} \leqslant$$

$$\leqslant \frac{|X| - |Y|}{|Y| \cdot (|X| - |Y|)} = \frac{1}{|Y|}.$$

Таким образом:

$$P(h(\widetilde{x}) = y) \leqslant \frac{1}{|Y|}.$$

2. Рассмотрим случай $(x, \widetilde{y})$. Так как $y \neq \widetilde{y}$, то $h(x) \neq \widetilde{y}$. Следовательно:

$$P(h(x) = \widetilde{y}) = 0.$$

3. Остался случай $(\widetilde{x}, \widetilde{y})$. Вычислим вероятность события, при котором $h(\widetilde{x}) = \widetilde{y}$. Пространством элементарных исходов будет такое множество: $X\backslash\{x\} \times Y\backslash\{y\}$.

Пусть, как и в предложении 10.1, B — матрица над множеством $\{0, 1\}$ размером $(|X| - 1) \times (|Y| - 1)$, в которой строки занумерованы элементами множества $X\backslash\{x\}$, а столбцы — элементами множества $Y\backslash\{y\}$. В каждой клетке с номером $(\widetilde{x}, \widetilde{y}) \in X\backslash\{x\} \times Y\backslash\{y\}$ поставим единицу, если $h(\widetilde{x}) = \widetilde{y}$, иначе 0. Очевидно, что в каждой строке данной матрицы будет не более одной единицы, поэтому количество пар $(\widetilde{x}, \widetilde{y})$, для которых $h(\widetilde{x}) = \widetilde{y}$, не превосходит $|X| - 1$. Следовательно:

$$P(h(\widetilde{x}) = \widetilde{y}) \leqslant \frac{|X| - 1}{(|X| - 1) \cdot (|Y| - 1)} = \frac{1}{|Y| - 1}.$$

Таким образом, если, например, $m = 256$ (длина свертки), то из рассмотренных выше трех случаев видно, что вероятность

не обнаружить изменения данных не превышает числа $\frac{1}{2^{256}-1}$, т.е. такая вероятность крайне мала.

Теперь рассмотрим такую ситуацию. Предположим, что противнику известна хеш-функция h (т.е. алгоритм нахождения значения y по аргументу x) и он перехватил сообщение (x, y), где $y = h(x)$. Противник может подменить (x, y) сообщением $(\widetilde{x}, \widetilde{y})$, где $\widetilde{y} = h(\widetilde{x})$. Тогда законный пользователь может не обнаружить подмены сообщения, так как проверка целостности данных пройдет успешно.

Чтобы избежать подмены или имитации сообщения, применяются ключевые хеш-функции:

$$h : X \times K \to Y.$$

В этом случае получить значение хеш-функции $y = h(x, k)$, где $x \in X$, $k \in K$, может только обладатель секретного ключа k. Поэтому подменить сообщение (x, y) так, чтобы это не было обнаружено, почти невозможно.

10.3. Криптографические хеш-функции

Определим основные требования, которым должна удовлетворять криптографическая хеш-функция, не зависящая от ключа.

- **Практическая эффективность**. Значение $h(x)$ должно легко вычисляться для любого $x \in X$.
- **Сложность вычисления прообразов**. Для произвольного значения $y \in Y$ задача нахождения такого $x \in X$, что $h(x) = y$, должна быть вычислительно трудоемкой (практически неразрешимой).
- **Устойчивость к коллизиям**. Должно быть вычислительно трудоемко найти какую-либо пару $x_1, x_2 \in X$, $x_1 \neq x_2$, что $h(x_1) = h(x_2)$.

- **Устойчивость к нахождению второго прообраза**. Должно быть вычислительно трудоемко по заданному $x \in X$ найти какое-либо значение $\widetilde{x} \in X$, $x \neq \widetilde{x}$, что $h(x) = h(\widetilde{x})$.

При этом свойства 1 и 2 эквивалентны тому, что функция h должна быть односторонней, а из свойства 3 следует свойство 4.

Предложения 10.2 и 10.3 показывают, что если хеш-функция является односторонней, то поиск коллизий представляет собой очень трудную задачу и ее сложность зависит от мощности множества образов данной хеш-функции.

Следующее предложение показывает, что если хеш-функция не является односторонней, то в этом случае поиск коллизий значительно упрощается.

Предложение 10.4. Пусть имеется некоторая сбалансированная функция $h : X \to Y$, где X, Y — конечные множества, причем выполнено неравенство $0 < |Y| < |X|$. Если имеется эффективный алгоритм f вычисления прообраза $x \in X$ по известному $y \in Y$, такого что $y = h(x)$, то имеется и эффективный вероятностный алгоритм нахождения коллизии для h с вероятностью $1 - \dfrac{|Y|}{|X|}$.

Доказательство. Зафиксируем произвольным образом $x \in X$ и вычислим значение $h(x) = y$. Так как имеется эффективный алгоритм f вычисления прообразов, то вычислим $\widetilde{x} = f(y)$. Элемент $\widetilde{x}$ принадлежит множеству $h^{-1}(y)$. При этом может так случиться, что $\widetilde{x} \neq x$. Вероятность события $\widetilde{x} \neq x$ равна такому числу:

$$\frac{|h^{-1}(y)| - 1}{|h^{-1}(y)|} = \frac{|X| - |Y|}{|X|}.$$

Теперь подсчитаем вероятность коллизии. Наш эксперимент заключается в случайном выборе элемента $x \in X$, вычислениях $y = h(x)$, $\widetilde{x} = f(y)$ и сравнении x и $\widetilde{x}$. Так как элемент $x \in X$ выбирается случайно, то пространством элементарных исходов

Ω будет являться множество X ($\Omega = X$) с равномерным распределением вероятностей:

$$P(x) = \frac{1}{|X|}, \quad x \in X.$$

Обозначим через B событие, заключающееся в том, что произошла коллизия. Тогда по формуле полной вероятности:

$$P(B) = \sum_{x \in X} P(x) P(B \mid x),$$

причем для любого $x \in X$

$$P(B \mid x) = \frac{|X| - |Y|}{|X|}.$$

Следовательно:

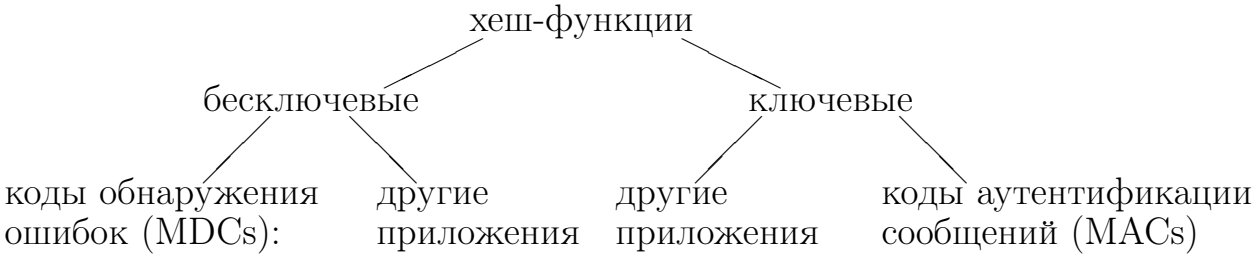
$$P(B) = \frac{1}{|X|} \cdot \sum_{x \in X} \frac{|X| - |Y|}{|X|} = \frac{|X| - |Y|}{|X|} = 1 - \frac{|Y|}{|X|}. \qquad \square$$

Следствие 10.1. Если в условии предложения 10.4 $\dfrac{|Y|}{|X|} \to 0$, то вероятность нахождения коллизии для h стремится к 1.

Криптографические хеш-функции применяются прежде всего для установления целостности данных и аутентификации источника данных. *Целостность данных* — свойство, позволяющее убедиться в том, что данные не были изменены неавторизованным способом с тех пор, как они были созданы авторизованным источником. *Аутентификации источника данных* — получение подтверждения того, что данное сообщение было получено именно от указанного источника.

В криптографии особо выделяются два типа криптографических хеш-функций — ключевые (задаваемые ключом) и бесключевые (не зависящие от ключа). Ключевые хеш-функции применяют в системах с симметричными ключами и доверяющими друг другу сторонами. Данные функции называют *кодами аутентификации сообщений* (message authentication code — MAC). Хеш-функции, не зависящие от ключа, называют *кодами обнаружения ошибок* (modification detection code — MDC).

Ниже представлена упрощенная классификация криптографических хеш-функций.



- односторонняя функция,
- устойчивость к коллизиям,
- устойчивость к нахождению 2-го прообраза

10.4. Построение хеш-функций

Одношаговые сжимающие отображения. Так как хеш-функция не накладывает ограничение на длину сообщения, то в большинстве случаев она строится на основе так называемых *одношаговых сжимающих отображений.*

Пусть f — некоторая функция:

$$f : V_n \times V_m \to V_m.$$

Разобьем исходное сообщение x на (двоичные) блоки длины n : $x = x_1 \dots x_l$, где все $x_i \in V_n$. Пусть H_0 — некоторый начальный фиксированный вектор из множества V_m. Положим:

$$H_i = f(x_i, H_{i-1}), \quad i = 1, 2, \dots, l. \tag{10.1}$$

После чего полагаем $h(x) = H_l$. Таким образом, построена хеш-функция на базе итеративного механизма.

Если функция f не зависит от ключа, то для исключения возможности перебора коротких сообщений вектор H_0 можно составить из фрагментов, указывающих дату, время и т.п.

Некоторые ключевые хеш-функции строятся на базе алгоритмов блочного шифрования E_k. Пусть, например, $x = x_1 \dots x_l$ — некоторое сообщение, $x_1, \dots, x_l \in V_n$, и $H_0 \in V_n$. Определим:

$$H_i = E_k(x_i \oplus H_{i-1}), \quad i = 1, 2, \dots, l,$$

$$h(x) = H_l.$$

Например, в алгоритме ГОСТ Р 34.12-2015 данный алгоритм называется выработкой имитовставки. В данном случае в качестве вектора H_0 можно взять нулевой вектор.

Построение хеш-функций, задаваемых ключом, на основе бесключевых. На основе бесключевых хеш-функций можно строить хеш-функции, задаваемые ключом. В этом случае при вычислении свертки к сообщению «подмешивают» значение ключа.

Заметим, что если ключ просто дописывать в начало или в конец сообщения, то это может привести к потенциальным слабостям, позволяющим в некоторых случаях осуществлять модификацию сообщений.

Предположим, что ключ k добавляют к началу сообщения $x : h_k(x) = h(k||x)$, где $||$ — операция конкатенации (приписывания) сообщений (в дальнейшем будем вместо операции конкатенации использовать запятую). Пусть функция h построена на основе одношаговой сжимающей функции по формуле (10.1). Если противник перехватит сообщение $(x, h(k, x))$, $x = x_1 \dots x_l$, то он может к сообщению x добавить сообщение $y = y_1 \dots y_m$ и вычислить:

$$H_{l+1} = f(y_1, H_l),\ H_{l+2} = f(y_2, H_{l+1}), \dots,\ H_{l+m} = f(y_m, H_{l+m-1}).$$

После этого противник может заменить сообщение (x, H_l) сообщением (x, y, H_{l+m}). При этом понятно, что $h_k(x, y) = H_{l+m}$.

Более предпочтительными являются способы «подмешивания» ключа, при которых ключ приписывают к сообщению не в одном, а в нескольких местах. Например [62]:

$$H = h(k, y, x, k),$$
$$H = h(k, y_1, h(k, y_2, x)),$$

где y, y_1, y_2 — дополнения ключа k до размера, кратного длине блока n.

Еще один известный пример построения хеш-функции, заданной ключом, на основе бесключевой хеш-функции дает алгоритм HMAC, согласно которому:

$$\text{HMAC}_k(x) = h(k^* \oplus a, h(k^* \oplus b, x)),$$

где k^* — дополнение ключа k нулями в начале до получения полного блока; a, b — некоторые константы, длина которых совпадает с размером блока [61].

Бесключевые хеш-функции на основе блочных шифров. Такие хеш-функции прежде всего ориентированы на обеспечение подлинности и целостности информации и предназначены для построения криптографических систем электронной подписи сообщений.

Пусть $y = E_k(x)$ — алгоритм блочного шифрования. Если размер ключа k и блоков x и y совпадает, то на основе блочного шифра E_k существует несколько различных бесключевых схем хеширования:

1) $H_i = E_{H_{i-1}}(x_i) \oplus x_i$;

2) $H_i = E_{x_i}(H_{i-1}) \oplus H_{i-1}$,

где $i = 1, 2, \ldots, l, \quad h(x) = H_l$. Заметим, что первое преобразование лежит в основе российского стандарта функции хеширования ГОСТ Р 34.11-94, а вторая — в основе американского стандарта SHA-1.

Если размеры блока и длина ключа не совпадают, то можно рассматривать следующие конструкции. Пусть n — размер блока, s — длина ключа, $G : V_n \to V_s$ — некоторое отображение. Ниже приведены одношаговые сжимающие функции, построенные на основе алгоритма E_k :

3) $H_i = E_{G(H_{i-1})}(x_i) \oplus x_i$ (Matyas-Meyer-Oseas);

4) $H_i = E_{G(H_{i-1})}(x_i) \oplus x_i \oplus H_{i-1}$ (Miyaguchi-Preneel).

10.5. Хеш-функция ГОСТ Р 34.11-2012

«Стрибог» — это семейство хеш-функций, включающее в себя две функции: функцию с длиной свертки в 256 бит и функцию с длиной свертки в 512 бит. Обе эти функции имеют одинаковую структуру и отличаются друг от друга только начальным значением вектора инициализации IV. Входными данными для обоих функций является блок данных длиной 512 бит. В случае, если длина сообщения больше 512 бит, то происходит разбиение

сообщения на блоки. В случае же, если длина меньше 512 бит, то производится дополнение сообщения.

Концепция этой криптографической хеш-функции состоит в использовании минимального числа элементов, но необходимого для того, чтобы обеспечить устойчивость ко всем известным криптографическим атакам на хеш-функцию. Принципы построения хеш-функции «Стрибог»:

- неприменимость известных атак;
- использование хорошо изученных конструкций и преобразований;
- обеспечение каждого структурного элемента конкретными свойствами;
- использование самого простого для анализа и реализации варианта (если их несколько);
- максимальная производительность программной реализации.

Функция сжатия. В хеш-функции важным элементом является функция сжатия. В ГОСТ Р 34.11-2012 функция сжатия основана на конструкции Миягучи-Пренела (Miyaguchi-Preneel):

$$\begin{gathered} g_N : V_{512} \times V_{512} \to V_{512}, \quad N \in V_{512}, \\ g_N(h, x) = E_{(L \circ P \circ S \circ X[N])(h)}(x) \oplus h \oplus x, \end{gathered} \tag{10.2}$$

где функции L, P, S, X представлены ниже.

XSPL-шифр. Блочный шифр E_k из схемы (10.2) является 13-раундовым алгоритмом с длинами ключа и входного блока 512 бит:

$$E_k(x) = (X[q_{13}] \circ \varphi_{q_{12}} \circ \ldots \circ \varphi_{q_1})(x) : V_{512} \to V_{512},$$

где $\varphi_{q_i}(x) : V_{512} \to V_{512}$ — раундовая функция, q_i — раундовый ключ. Раундовая функция φ_q состоит из следующих слоев.

X-преобразование. На вход функции X подаются две последовательности длиной 512 бит каждая, выходом функции является XOR (сложение по модулю два) этих последовательностей:

$$X[k] : V_{512} \to V_{512}, \quad X[k](a) = k \oplus a, \ k, a \in V_{512}.$$

S-преобразование. Функция S является подстановкой. Каждый байт из 512-битной входной последовательности заменяется соответствующим байтом из таблицы подстановок π :

$$S : V_{512} \to V_{512}, \quad S(a) = S(a_{63}|| \dots ||a_0) = \pi(a_{63})|| \dots ||\pi(a_0),$$

где $a = a_{63}|| \dots ||a_0 \in V_{512}$, $a_i \in V_8$, $i = 0, \dots 63$, подстановка $\pi : V_8 \to V_8$ является константой, определенной в стандарте. Данная подстановка совпадает с аналогичной подстановкой в шифре «Кузнечик» из ГОСТ Р 34.12-2015 (см. параграф 8.8).

P-преобразование. Функция перестановки байтов во входной последовательности:

$$P : V_{512} \to V_{512}, \quad P(a) = P(a_{63}|| \dots ||a_0) = a_{\tau(63)}|| \dots ||a_{\tau(0)},$$

где τ — константная перестановка, заданная на 64-элементном множестве $\{0, 1, \dots 63\}$.

L-преобразование. Представляет собой умножение 64-битного входного вектора на двоичную константную матрицу A размера 64×64 над полем $GF(2)$:

$$L : V_{512} \to V_{512}, \quad L(a) = L(a_7|| \dots ||a_0) = l(a_7)|| \dots ||l(a_0),$$

где $a = a_7|| \dots ||a_0 \in V_{512}$, $a_i \in V_{64}$, $i = 0, \dots 7$,

$$l(x) = l(x_{63} \dots x_0) = (x_{63} \dots x_0) \cdot A, \quad x \in V_{64}.$$

Итак, раундовая функция φ_q имеет такой вид:

$$\varphi_q(x) = (L \circ P \circ S \circ X[q])(x).$$

Раундовые ключи $q_1, \dots, q_{13}$ вычисляются следующим образом:

$$q_1 = k = (L \circ P \circ S \circ X[N])(h),$$

$$q_i = (L \circ P \circ S \circ X[C_{i-1}])(q_{i-1}), \quad i = 2, \dots, 13,$$

где C_i $(i = 1, \dots, 12)$ — константы, заданные в ГОСТе.

Алгоритм вычисления хеш-функции

1. Присвоить начальные значения текущих величин:

   $IV := 0^{512}$ для хеш-функции с длиной свертки 512 бит.

$IV := (00000001)^{64}$ для хеш-функции с длиной свертки 256 бит, где IV — вектор инициализации,

$h := IV$, $N := 0^{512}$, $\Sigma := 0^{512}$.

2. Проверить длину двоичного сообщения x. Если выполнено $0 \leqslant |x| < 512$ ($|x|$ — длина сообщения x), то перейти к пункту 3. В противном случае выполняем следующую последовательность вычислений.

   В переменную m записываем 512 последних бит сообщения x.

   $h := g_N(h, m)$.

   $N := (N + 512) \bmod 2^{512}$.

   $\Sigma = (\Sigma + m) \bmod 2^{512}$.

   Отбрасываем последние 512 бит из сообщения x и переходит к шагу 2.

3. $x := 0^{512-|x|}||1||x$ (вставляем в начало сообщения вектор $0\ldots01$ до длины 512).

   $h := g_N(h, x)$.

   $N := (N + |x|) \bmod 2^{512}$.

   $\Sigma := (\Sigma + x) \bmod 2^{512}$.

   $h := g_0(h, N)$.

   $h := g_0(h, \Sigma)$.

   Для хеш-функции с длиной свертки 256 бит в h записываем 256 старших бит переменной h.

Значение величины h, полученное в конце шага 3 является значением хеш-функции $h(x)$.

10.6. Парадокс дней рождений

Заметим, что трудоемкость подбора прообраза для однонаправленной функции или трудоемкость поиска второго прообраза оценивают величиной $O(2^n)$, в то же время трудоемкость

поиска коллизии оценивают величиной $O(2^{n/2})$, так как в данном случае применима атака, основанная на комбинаторном парадоксе дней рождений.

Парадокс дней рождений заключается в том, что вероятность p наличия коллизии в выборке из m элементом объема $O(\sqrt{m})$ принимает достаточно большое значение.

Далее нам понадобятся некоторые сведения из математического анализа.

Пусть функция f определена на некотором интервале (a, b) и пусть $a < x_1 < x_2 < b$. Через точки $(x_1, f(x_1))$ и $(x_2, f(x_2))$ проведем прямую $l(x)$, уравнение которой имеет такой вид:

$$l(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x - x_1). \tag{10.3}$$

Лемма 10.1. Для функции f, определенной на некотором интервале (a, b), следующие условия эквивалентны.

(i) Для любых точек x, x_1, x_2, $a < x_1 \leqslant x \leqslant x_2 < b$, выполнено неравенство $f(x) \leqslant l(x)$.

(ii) Для любого $\alpha \in [0, 1]$ и любых $x_1, x_2 \in (a, b)$ выполнено неравенство:

$$f((1 - \alpha)x_1 + \alpha x_2) \leqslant (1 - \alpha)f(x_1) + \alpha f(x_2).$$

(iii) Для любых $x_1, \ldots, x_n \in (a, b)$ и любых $\alpha_1, \ldots, \alpha_n \in [0, 1]$, $\alpha_1 + \ldots + \alpha_n = 1$, выполнено неравенство Йенсена:

$$f(\alpha_1 x_1 + \ldots + \alpha_n x_n) \leqslant \alpha_1 f(x_1) + \ldots + \alpha_n f(x_n).$$

Доказательство. Пусть выполнено условие (i). Рассмотрим отображение:

$$x = x(\alpha) = x_1 + \alpha(x_2 - x_1) = (1 - \alpha)x_1 + \alpha x_2.$$

Очевидно, что образом отрезка $[0, 1]$ при отображении $x(\alpha)$ является отрезок $[x_1, x_2]$: $x([0, 1]) = [x_1, x_2]$. Поэтому из неравенства $f(x) \leqslant l(x)$ будет следовать для любого $\alpha \in [0, 1]$ такое неравенство:

$$f((1-\alpha)x_1+\alpha x_2) \leqslant l((1-\alpha)x_1+\alpha x_2) \overset{(10.3)}{=} (1-\alpha)f(x_1)+\alpha f(x_2).$$

Поэтому из условия (i) следует (ii).

Пусть выполнено условие (ii). Как и ранее, рассмотрим отображение $x(\alpha) = (1-\alpha)x_1 + \alpha x_2$. Зафиксируем произвольное значение $x \in [x_1, x_2] \subset (a, b)$. Тогда найдется единственное $\alpha \in [0, 1]$, такое что $x = (1-\alpha)x_1 + \alpha x_2$. Выразив в последнем равенстве α и подставив его в неравенство:

$$f((1-\alpha)x_1 + \alpha x_2) \leqslant (1-\alpha)f(x_1) + \alpha f(x_2),$$

получим справедливость условия (i), поэтому из (ii) следует (i).

Покажем методом математической индукции по n, что из (ii) следует (iii) с очевидным основанием $n = 2$. Пусть:

$$x_1, \ldots, x_n \in (a, b), \quad \alpha_1, \ldots, \alpha_n \in [0, 1], \quad \alpha_1 + \ldots + \alpha_n = 1$$

и пусть, для определенности, $\alpha_1 > 0$. Тогда:

$$\beta = \alpha_1 + \ldots + \alpha_{n-1} > 0, \quad \frac{\alpha_1}{\beta} + \ldots + \frac{\alpha_{n-1}}{\beta} = 1.$$

Учитывая предположение и базу индукции, получаем:

$$\begin{gathered} f(\alpha_1 x_1 + \ldots + \alpha_n x_n) = \\ = f\left(\beta\left(\frac{\alpha_1}{\beta}x_1 + \ldots + \frac{\alpha_{n-1}}{\beta}x_{n-1}\right) + \alpha_n x_n\right) \leqslant \\ \leqslant \beta f\left(\frac{\alpha_1}{\beta}x_1 + \ldots + \frac{\alpha_{n-1}}{\beta}x_{n-1}\right) + \alpha_n f(x_n) \leqslant \\ \leqslant \alpha_1 f(x_1) + \ldots + \alpha_n f(x_n). \end{gathered}$$

При этом в первом неравенстве данной цепочки использовался тот факт, что:

$$\frac{\alpha_1}{\beta}x_1 + \ldots + \frac{\alpha_{n-1}}{\beta}x_{n-1} \in (a, b), \text{ так как } \frac{\alpha_1}{\beta} + \ldots + \frac{\alpha_{n-1}}{\beta} = 1.$$

Очевидно, что из условия (iii) следует (ii). □

Функция f, определенная на некотором интервале (a, b), называется *выпуклой вниз*, если выполнено одно из условий леммы 10.1. Заметим, что из леммы 10.1 для любого $k \geqslant 2$ следует такое неравенство:

$$\sum_{i=1}^{n} p_i^k \geqslant \frac{1}{n^{k-1}} \tag{10.4}$$

для произвольного набора неотрицательных действительных чисел $p_1, \ldots, p_n$, где $p_1 + \ldots + p_n = 1$. Действительно, функция x^k является выпуклой вниз на полуинтервале $[0, +\infty)$, поэтому:

$$(\alpha_1 p_1 + \ldots + \alpha_n p_n)^k \leqslant \alpha_1 p_1^k + \ldots + \alpha_n p_n^k$$

для любого набора $\alpha_1, \ldots, \alpha_n \in [0, 1]$, $\alpha_1 + \ldots + \alpha_n = 1$. Подставим в последнее неравенство такие значения: $\alpha_1 = \ldots = \alpha_n = \dfrac{1}{n}$. Получаем:

$$\frac{1}{n} \cdot \sum_{i=1}^{n} p_i^k \geqslant \left(\frac{1}{n} \cdot \sum_{i=1}^{n} p_i \right)^k = \frac{1}{n^k}.$$

Лемма 10.2. Пусть $p_1, \ldots, p_n$ — произвольный набор действительных чисел из отрезка $[0, 1]$ с условием, что $p_1 + \ldots + p_n = 1$, и пусть m — некоторое натуральное число, $m < n$. Тогда:

$$\sum_{i=1}^{n} p_i (1 - p_i)^m \leqslant \left(1 - \frac{1}{n} \right)^m.$$

Обозначим через $\mathcal{A}_n^k$ множество всех размещений без повторений из n элементов по k множества $\{1, \ldots, n\}$, т.е. множество $\mathcal{A}_n^k$ состоит из всех упорядоченных выборок вида $(i_1, \ldots, i_k)$, где:

$$\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$$

и $i_1, \ldots, i_k$ попарно различны.

Лемма 10.3. Пусть $p_1, \ldots, p_n$ —произвольный набор действительных чисел из отрезка $[0, 1]$ с условием, что $p_1 + \ldots + p_n = 1$. Тогда для любого фиксированного $2 \leqslant k \leqslant n$ будет выполнено такое неравенство:

$$\sum_{(i_1, \ldots, i_k) \in \mathcal{A}_n^k} p_{i_1} \ldots p_{i_k} \leqslant \prod_{i=1}^{k-1} \left(1 - \frac{i}{n} \right).$$

Доказательство. I. Рассмотрим сначала случай $0 < p_i < 1$, $i = 1, \ldots, n$.

1 способ доказательства проведем методом математической индукции. Пусть $k = 2$. Тогда:

$$\sum_{(i,j)\in\mathcal{A}_n^2} p_i p_j = \sum_{i=1}^{n} p_i \sum_{\substack{1\leqslant j\leqslant n \\ i\neq j}} p_j = \sum_{i=1}^{n} p_i(1-p_i) = 1-\sum_{i=1}^{n} p_i^2 \overset{(10.4)}{\leqslant} 1-\frac{1}{n}.$$

Предположим, что утверждение леммы верно для всех значений $2, \ldots, k-1$ при $k > 2$. Обозначим:

$$S^k(p_1, \ldots, p_n) = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1} \ldots p_{i_k}.$$

Тогда:

$$S^k(p_1, \ldots, p_n) = p_1 S^{k-1}(p_2, p_3, \ldots, p_n) +$$
$$+p_2 S^{k-1}(p_1, p_3, \ldots, p_n) + \ldots + p_n S^{k-1}(p_1, p_2, \ldots, p_{n-1}) =$$
$$= \sum_{i=1}^{n} p_i S^{k-1}(p_1, \ldots, \widehat{p_i}, \ldots, p_n) =$$
$$= \sum_{i=1}^{n} p_i (1-p_i)^{k-1} S^{k-1}\left(\frac{p_1}{1-p_i}, \ldots, \widehat{\frac{p_i}{1-p_i}}, \ldots, \frac{p_n}{1-p_i}\right).$$

Так как:

$$\frac{p_1}{1-p_i} + \ldots + \widehat{\frac{p_i}{1-p_i}} + \ldots + \frac{p_n}{1-p_i} = 1,$$

то по предположению индукции:

$$S^{k-1}\left(\frac{p_1}{1-p_i}, \ldots, \widehat{\frac{p_i}{1-p_i}}, \ldots, \frac{p_n}{1-p_i}\right) \leqslant \prod_{i=1}^{k-2}\left(1-\frac{i}{n-1}\right).$$

Поэтому, с учетом леммы 10.2:

$$S^k(p_1, \ldots, p_n) \leqslant \left(\sum_{i=1}^{n} p_i(1-p_i)^{k-1}\right) \prod_{i=1}^{k-2}\left(1-\frac{i}{n-1}\right) \leqslant$$
$$\leqslant \left(1-\frac{1}{n}\right)^{k-1} \prod_{i=1}^{k-2}\left(1-\frac{i}{n-1}\right) = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right).$$

2 способ доказательства проведем методом множителей Лагранжа. Рассмотрим нашу сумму:

$$f(p_1,\ldots,p_n)=\sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1}\ldots p_{i_k}=k!\left(\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} p_{i_1}\ldots p_{i_k}\right)$$

как функцию, определенную на открытом множестве:

$$G=\{(p_1,\ldots p_n)\in\mathbb{R}^n \mid 0<p_i<1,\ i=1,\ldots,n\}.$$

Найдем условные экстремумы функции $f(p_1,\ldots,p_n)$ на множестве G при условии, что $p_1+\ldots+p_n=1$. Функция Лагранжа в данном случае будет иметь такой вид:

$$F(p_1,\ldots,p_n,\lambda)=k!\left(\sum_{1\leqslant i_1<\ldots<i_k\leqslant n} p_{i_1}\ldots p_{i_k}\right)-\lambda\left(\sum_{i=1}^{n}p_i-1\right).$$

Частные производные функции F по переменным $p_1,\ldots,p_n,\lambda$ примут такой вид:

$$\frac{\partial F}{\partial p_j}=k!\left(\sum_{\substack{i_1,\ldots,i_{k-1}\in\{1,\ldots,n\}\setminus\{j\}\\ i_1<\ldots<i_{k-1}}} p_{i_1}\ldots p_{i_{k-1}}\right)+\lambda,\quad j=1,\ldots,n,$$

$$\frac{\partial F}{\partial\lambda}=\sum_{i=1}^{n}p_i-1.$$

Теперь требуется решить систему из $n+1$ уравнений относительно $p_1,\ldots,p_n,\lambda$:

$$\begin{cases}\dfrac{\partial F}{\partial p_j} &= 0,\ \ j=1,\ldots,n,\\ \dfrac{\partial F}{\partial\lambda} &= 0.\end{cases}\tag{10.5}$$

Заметим, что:

$$\frac{\partial F}{\partial p_1}-\frac{\partial F}{\partial p_2}=k!(p_2-p_1)\left(\sum_{\substack{i_1,\ldots,i_{k-2}\in\{1,\ldots,n\}\setminus\{1,2\}\\ i_1<\ldots<i_{k-2}}} p_{i_1}\ldots p_{i_{k-2}}\right),$$

$$\dots$$

$$\frac{\partial F}{\partial p_1} - \frac{\partial F}{\partial p_n} = k!(p_n - p_1) \left(\sum_{\substack{i_1,\dots,i_{k-2} \in \{1,\dots,n\} \setminus \{1,n\} \\ i_1 < \dots < i_{k-2}}} p_{i_1} \dots p_{i_{k-2}} \right).$$

Так как большие скобки данных равенств строго больше нуля (так как все $p_i > 0$), то из равенства

$$\frac{\partial F}{\partial p_1} - \frac{\partial F}{\partial p_j} = 0, \quad j = 2, \dots, n,$$

следует, что $p_1 = p_2 = \dots = p_n = 1/n$. Нетрудно видеть, что решением системы (10.5), и притом единственным, являются значения:

$$p_i = \frac{1}{n}, \ i = 1, \dots, n, \ \lambda = \prod_{i=1}^{k-1} \left(\frac{i}{n} - 1 \right).$$

Дальнейшие исследования показывают, что при данных значениях функция $f(p_1, \dots, p_n)$ в области G достигает максимума при условии $p_1 + \dots + p_n = 1$. Поэтому:

$$f(p_1, \dots, p_n) = k! \left(\sum_{1 \leqslant i_1 < \dots < i_k \leqslant n} p_{i_1} \dots p_{i_k} \right) \leqslant$$

$$\leqslant k! \cdot C_n^k = \prod_{i=1}^{k-1} \left(1 - \frac{i}{n} \right).$$

II. Рассмотрим более общий случай, когда $0 \leqslant p_i \leqslant 1$, $i = 1, \dots, n$, $p_1 + \dots + p_n = 1$. Без ограничения общности можно считать, что:

$$p_1, \dots, p_m > 0, \quad p_{m+1} = \dots = p_n = 0,$$

$$1 \leqslant m \leqslant n, \quad p_1 + \dots + p_m = 1.$$

Если $k > m$, то:

$$\sum_{(i_1,\dots,i_k) \in \mathcal{A}_n^k} p_{i_1} \dots p_{i_k} = 0 < \prod_{i=1}^{k-1} \left(1 - \frac{i}{n} \right).$$

Если же $k \leqslant m$, то из **I** следует, что:

$$\sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1}\ldots p_{i_k} = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_m^k} p_{i_1}\ldots p_{i_k} \leqslant$$

$$\leqslant \prod_{i=1}^{k-1}\left(1-\frac{i}{m}\right) \leqslant \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right). \qquad \square$$

Лемма 10.4. Пусть имеется полная группа событий $A_1,\ldots,$ A_n. В каждом независимом испытании может произойти одно и только одно из данных событий. Проводится k независимых испытаний, $k < n$. Пусть B_k — событие, заключающееся в том, что в k независимых испытаниях хотя бы одно из событий $A_1,\ldots,$ A_n произойдет более одного раза. Тогда для вероятности $P(B_k)$ события B_k выполнено следующее неравенство:

$$P(B_k) > 1 - e^{-\frac{(k-1)^2}{2n}}.$$

Доказательство. Пусть $p_1,\ldots,$ p_n — вероятности соответствующих событий $A_1,\ldots,$ A_n, $\overline{B}_k$ — противоположное событие к B_k. Понятно, что:

$$P(\overline{B}_k) = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1}\ldots p_{i_k}.$$

Если $p_1 = \ldots = p_k = \dfrac{1}{n}$, то:

$$P(\overline{B}_k) = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1}\ldots p_{i_k} = \frac{n!}{(n-k)!}\cdot\frac{1}{n^k} = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right).$$

В противном случае, учитывая лемму 10.3, имеем:

$$P(\overline{B}_k) = \sum_{(i_1,\ldots,i_k)\in\mathcal{A}_n^k} p_{i_1}\ldots p_{i_k} \leqslant \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right).$$

Учитывая, что при $0 < x < 1$:

$$\ln(1-x) = -x - \frac{x^2}{2} - \ldots - \frac{x^n}{n} - \ldots < -x,$$

получаем такую цепочку неравенств:

$$\ln P(\overline{B}_k) \leqslant \ln \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right) = \sum_{i=1}^{k-1} \ln\left(1-\frac{i}{n}\right) < -\sum_{i=1}^{k-1}\frac{i}{n} =$$

$$= -\frac{(k-1)k}{2n} < -\frac{(k-1)^2}{2n}.$$

Так как $P(\overline{B}_k) = 1 - P(B_m)$, то $P(B_k) > 1 - e^{-\frac{(k-1)^2}{2n}}$. □

Предложение 10.5. Пусть $f : X \to Y$ — некоторое сюръективное отображение, где X, Y — конечные множества. Случайным образом выбираются с возвратом элементы $x_1, \ldots, x_k \in X$. Тогда для вероятности p того, что среди данных k элементов найдутся хотя бы два элемента x_i и x_j $(i \neq j)$ со свойством $f(x_i) = f(x_j)$, выполнено следующее неравенство:

$$p > 1 - e^{-\frac{(k-1)^2}{2|Y|}}.$$

Доказательство. Пусть $Y = \{y_1, \ldots, y_n\}$, A_i — событие, заключающееся в том, что при случайном выборе $x \in X$ будет выполнено равенство $f(x) = y_i$ $(i = 1, \ldots, n)$. Тогда для оценки вероятности p остается применить лемму 10.4. □

В предложении 10.5 учитываются и варианты $x_i = x_j$ при $i \neq j$. При оценке вероятности коллизии такие варианты необходимо исключить. Итак, пусть имеется некоторая сбалансированная функция $h : X \to Y$, где X, Y — конечные множества. Будем считать, что $|X| \gg |Y|$ и $\dfrac{|X|}{|Y|} \gg k$. Тогда в этом случае для функции h можно применить предложение 10.5 с оговоркой, что $x_i = x_j$ при $i \neq j$: *для вероятности коллизии p в выборке $x_1, \ldots, x_k \in X$ справедлива оценка* $p > 1 - e^{-\frac{(k-1)^2}{2|Y|}}$.

Пусть p — некоторое действительное число, $0 < p < 1$. Зададимся таким вопросом: сколько необходимо выбрать случайным образом элементов из множества X, чтобы с вероятностью не менее p среди них нашлись два таких элемента x_1 и x_2, $x_1 \neq x_2$, что $h(x_1) = h(x_2)$. Из сказанного выше следует такое

Предложение 10.6. Пусть p — некоторое действительное число, $0 < p < 1$. Пусть также имеется некоторая сбалансированная функция $h : X \to Y$, где X, Y — конечные множества, и $|X| \gg |Y|$. Случайным образом выбираются k различных элементов из множества X. Для того чтобы среди данных k элементов с вероятностью не менее p нашлись два элемента x_1 и x_2, $x_1 \neq x_2$, со свойством $h(x_1) = h(x_2)$, достаточно, чтобы число k удовлетворяло следующему неравенству:

$$k > \sqrt{2|Y| \ln \frac{1}{1-p}} + 1. \tag{10.6}$$

Если обозначить

$$c_p = \sqrt{2 \ln \frac{1}{1-p}},$$

то для числа k неравенство можно записать в таком виде:

$$k > c_p \sqrt{|Y|} + 1.$$

Например, если $p = 1/2$, то $c_p \approx 1.17741$.

В качестве практического применения предложения 10.6 приведем такую комбинаторную задачу, известную как *парадокс дней рождений*, которая формулируется следующим образом: какой должна быть численность группы случайно выбранных людей, чтобы среди них с вероятностью $1/2$ нашлись бы два человека с одинаковым днем рождения.

Обозначим через X множество всех людей на земле, а через Y — все даты, начиная с 1 января и заканчивая 31 декабря. Функция $h : X \to Y$ сопоставляет каждому человеку его день рождения. Будем считать, что h является сбалансированным отображением. Используя неравенство (10.6), получаем, что $k > 23$. Таким образом, достаточно случайным образом выбрать группу из 24 человек, чтобы среди них с вероятностью $1/2$ нашлись бы два человека с одинаковым днем рождения.

Применяя предложение 10.6 к хеш-функции, получаем, что для поиска коллизии функции $h : X \to Y$ при $|Y| = 2^m$ при вероятности успеха p потребуется просмотреть $k > c_p \sqrt{2^m}$ элементов множества X.

Глава 11. Коды аутентификации

11.1. Основные понятия

Рассмотрим более подробно ключевые хеш-функции, которые очень часто применяются для обеспечения целостности данных и аутентификации источника данных и носят название кодов аутентификации.

Код аутентичности сообщения — в протоколах аутентификации сообщений с доверяющими друг другу участниками — специальный набор символов, добавляемый к сообщению и предназначенный для обеспечения его целостности и аутентификации источника данных.

Как правило, код аутентификации сопоставляет сообщению его код аутентичности сообщения. Алгоритм принятия решения о подлинности информации основан на проверке значения кода аутентичности сообщения. К коду аутентификации предъявляются следующие требования:

- невозможность вычисления кода аутентичности для заданного сообщения без знания ключа;
- невозможность подбора для одного или нескольких сообщений с известными значениями кода аутентичности другого сообщения с известным значением кода аутентичности без знания ключа.

Нас будет интересовать математическое определение данного понятия.

Определение 11.1. *Кодом аутентификации* (без сокрытия) называется четверка (X, K, Y, h), где X — конечное множество сообщений, K — конечное множество ключей, Y — конечное

множество сверток, h — ключевая хеш-функция и выполнено равенство

$$Y = \bigcup_{k \in K} h_k(X).$$

Пусть $P(X)$ и $P(K)$ — априорные распределения вероятностей соответственно на множествах X и K, не содержащие нулевых вероятностей. Понятно, что распределения $P(X)$ и $P(K)$ естественным образом индуцируют распределение вероятностей $P(Y)$:

$$P_Y(y) = \sum_{\substack{(x,k) \in X \times K, \\ h_k(x) = y}} P_X(x) \cdot P_K(k).$$

Заметим, что потенциальный противник может осуществлять не только пассивные действия относительно передаваемых по каналу связи сообщений, которые заключаются, например, в подслушивании или перехвате сообщений, но также и активные атаки, заключающиеся в *имитации* или *подмене* сообщений.

Пусть канал связи готов к работе и на приеме установлены действующие ключи $k \in K$, но в данный момент времени никакого сообщения вида (x, y), где $x \in X$, $y = h_k(x)$, не передается. Тогда в этом случае противником может быть предпринята попытка имитации сообщения парой $(\widetilde{x}, \widetilde{y}) \in X \times Y$.

Рассмотрим вероятностное пространство $(\Omega = K, F_K, P_K)$. Зафиксируем $(x, y) \in X \times Y$. Обозначим через $K(x, y)$ следующее множество:

$$K(x, y) = \{k \in K \mid h_k(x) = y\}.$$

Под обозначением $K(x, y)$ будем также понимать событие из алгебры событий F_K, заключающееся в том, что при случайном выборе ключа $k \in K$ будет выполнено равенство $h_k(x) = y$. Тогда событию $K(x, y)$ будут благоприятствовать все элементы из множества $K(x, y)$ и только они. Поэтому:

$$P(K(x, y)) = \sum_{k \in K(x,y)} P_K(k).$$

Поскольку противник имеет возможность выбора $(x, y) \in X \times Y$, его шансы на успех при имитации сообщения выражаются

такой величиной:

$$P_{im} = \max_{(x,y)\in X\times Y} P(K(x,y)).$$

Если же в данный момент передается некоторое сообщение

$$(x,y) \in X \times Y,\ y = h_k(x),$$

то противник может заменить его на $(\widetilde{x},\widetilde{y}) \in X\times Y$, $\widetilde{x} \neq x$. При этом он будет рассчитывать на то, что на действующем ключе k при проверке будет выполнено равенство $\widetilde{y} = h_k(\widetilde{x})$. Чем больше вероятность этого события, тем успешнее будет попытка подмены. Пусть "$K(\widetilde{x},\widetilde{y}) \mid K(x,y)$" — событие, заключающееся в попытке подмены сообщения (x,y) сообщением $(\widetilde{x},\widetilde{y})$. Применяя теорему о произведении вероятностей, получаем, что:

$$P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)) = \frac{P(K(x,y)\ \cap\ K(\widetilde{x},\widetilde{y}))}{P(K(x,y))}.$$

Тогда вероятность подмены сообщения будет вычисляться по следующей формуле:

$$P_{podm} = \max_{\substack{x,\widetilde{x}\in X,\ y,\widetilde{y}\in Y\\ x\neq\widetilde{x}}} P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)).$$

Теорема 11.1. Для любого кода аутентификации (X,K,Y,h) справедливы следующие утверждения:

(i)

$$P_{im} \geqslant \frac{1}{|Y|}, \tag{11.1}$$

причем равенство в (11.1) достигается тогда и только тогда, когда для всех $(x,y) \in X\times Y$ выполнено равенство $P(K(x,y)) = \dfrac{1}{|Y|}$.

(ii)

$$P_{podm} \geqslant \frac{1}{|Y|}, \tag{11.2}$$

причем равенство в (11.2) имеет место тогда и только тогда, когда для любых $x,\widetilde{x} \in X$, $y,\widetilde{y} \in Y$, $x \neq \widetilde{x}$, выполнено равенство $P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)) = \dfrac{1}{|Y|}$.

(*iii*) P_{im} и P_{podm} одновременно достигают нижней границы $\left(P_{im} = P_{podm} = \dfrac{1}{|Y|}\right)$ тогда и только тогда, когда для любых $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ выполнено равенство

$$P(K(x,y) \cap K(\widetilde{x},\widetilde{y})) = \frac{1}{|Y|^2}.$$

Доказательство. (*i*) Зафиксируем произвольное $x \in X$. Тогда:

$$1 = \sum_{y \in Y} P(K(x,y)) \leqslant \sum_{y \in Y} \max_{(x,y) \in X \times Y} P(K(x,y)) = P_{im} \cdot |Y|,$$

что доказывает неравенство $P_{im} \geqslant \dfrac{1}{|Y|}$.

Далее, если для любых $(x,y) \in X \times Y$ выполнено равенство $P(K(x,y)) = \dfrac{1}{|Y|}$, то по определению P_{im} следует, что $P_{im} = \dfrac{1}{|Y|}$.

Обратно, пусть $P_{im} = \dfrac{1}{|Y|}$. Предположим, что для некоторого $(x_0, y_0) \in X \times Y$ выполнено строгое неравенство $P(K(x,y)) < \dfrac{1}{|Y|}$. Тогда:

$$1 = \sum_{y \in Y} P(K(x_0,y)) = P(K(x_0,y_0)) + \sum_{y \in Y \setminus \{y_0\}} P(K(x_0,y)) <$$

$$< P_{im} + (|Y| - 1) \cdot P_{im} = |Y| \cdot P_{im}.$$

Противоречие.

(*ii*) Зафиксируем произвольным образом элементы $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y \in Y$. Тогда:

$$1 = \sum_{\widetilde{y} \in Y} P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)) \leqslant \sum_{\widetilde{y} \in Y} P_{podm} = |Y| \cdot P_{podm},$$

что доказывает требуемое неравенство.

Понятно, что если для любых $x, \widetilde{x} \in X$, $y, \widetilde{y} \in Y$, $x \neq \widetilde{x}$, выполнено равенство $P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)) = \dfrac{1}{|Y|}$, то $P_{podm} = \dfrac{1}{|Y|}$.

Обратно, пусть $P_{podm} = \dfrac{1}{|Y|}$. Предположим, что найдутся такие $x_1, x_2 \in X$, $y_1, y_2 \in Y$, $x_1 \neq x_2$, что:

$$P(K(x_1, y_1) \mid K(x_2, y_2)) < \frac{1}{|Y|}.$$

Тогда:

$$1 = \sum_{y \in Y} P(K(x_1, y) \mid K(x_2, y_2)) =$$

$$= P(K(x_1, y_1) \mid K(x_2, y_2)) + \sum_{y \in Y \setminus \{y_1\}} P(K(x_1, y) \mid K(x_2, y_2)) <$$

$$< P_{podm} + (|Y| - 1) \cdot P_{podm} = |Y| \cdot P_{podm}.$$

Противоречие.

(iii) Пусть $P_{im} = P_{podm} = \dfrac{1}{|Y|}$. Тогда из (i) и (ii) следует, что для любых $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ выполнены равенства:

$$P(K(x, y)) = \frac{1}{|Y|}, \quad P(K(\widetilde{x}, \widetilde{y}) \mid K(x, y)) = \frac{1}{|Y|}.$$

Поэтому:

$$P(K(x, y) \cap K(\widetilde{x}, \widetilde{y})) = P(K(x, y)) \cdot P(K(\widetilde{x}, \widetilde{y}) \mid K(x, y)) = \frac{1}{|Y|^2}.$$

Обратно, пусть для любых $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ выполнено равенство:

$$P(K(x, y) \ \cap \ K(\widetilde{x}, \widetilde{y})) = \frac{1}{|Y|^2}.$$

По формуле полной вероятности имеем:

$$P(K(\widetilde{x}, \widetilde{y})) = \sum_{y \in Y} P(K(x, y)) \cdot P(K(\widetilde{x}, \widetilde{y}) \mid K(x, y)) =$$

$$= \sum_{y \in Y} P(K(x, y) \ \cap \ K(\widetilde{x}, \widetilde{y})) = \frac{1}{|Y|}.$$

Поэтому из последних равенств следует, что:

$$P(K(\widetilde{x}, \widetilde{y}) \mid K(x, y)) = \frac{1}{|Y|}.$$

Осталось применить пункты (i) и (ii). $\square$

Предложение 11.1. Если для кода аутентификации выполнены неравенства $|X| > 1$ и $|K| < 2|Y|$, то $P_{podm} = 1$.

Доказательство. Так как $|K| < 2|Y|$, то найдутся такие элементы $x \in X$, $y \in Y$, что $|K(x, y)| = 1$. Пусть $k \in K(x, y)$. Поскольку $|X| > 1$, то зафиксируем произвольным образом элемент $\widetilde{x} \in X \backslash \{x\}$ и вычислим $\widetilde{y} = h_k(\widetilde{x})$. При этом

$$P(K(\widetilde{x}, \widetilde{y}) \mid K(x, y)) = 1,$$

поэтому $P_{podm} = 1$. □

11.2. Оптимальные коды аутентификации

Код аутентификации со свойством $P_{im} = P_{podm} = 1/|Y|$ называется *оптимальным*.

Ортогональные массивы. Для построения оптимальных кодов аутентификации понадобится понятие ортогональной таблицы.

Определение 11.2. *Ортогональной таблицей* $OA(s, n)$ над s-элементным множеством $Y = \{y_1, \ldots, y_s\}$ называется матрица порядка $s^2 \times n$ над множеством Y с тем условием, что для любых двух столбцов данной матрицы каждая из пар $(y_i, y_j) \in Y \times Y$ встречается ровно один раз.

Пример 11.1. Ортогональная таблица $OA(2, 3)$ над множеством $\{0, 1\}$:

0	0	0
0	1	1
1	0	1
1	1	0

Лемма 11.1. Для ортогональной таблицы $OA(s, n)$ над множеством Y, $|Y| = s$, справедливы следующие условия:

(i) если в $OA(s, n)$ вычеркнуть t столбцов $(0 \leqslant t \leqslant n - 2)$, то получится ортогональная таблица $OA(s, n - t)$;

(ii) перестановка произвольных строк (столбцов) ортогональной таблицы $OA(s, n)$ дает ортогональную таблицу с теми же параметрами;

(iii) если в некотором столбце $OA(s,n)$ применить некоторую подстановку над элементами множества Y, то получится ортогональная таблица с теми же параметрами.

Предложение 11.2. Для произвольной ортогональной таблицы $OA(s,n)$ над множеством $Y=\{y_1,\ldots,y_s\}$ выполнено следующее неравенство: $n\leqslant s+1$.

Доказательство. В соответствии со свойством (iii) леммы 11.1 переставим элементы в столбцах таблицы $OA(s,n)$ так, чтобы первая строка полученной ортогональной таблицы состояла только из элемента y_1. Тогда в оставшихся строках данной таблицы может присутствовать не более одного элемента y_1.

При этом в каждом столбце произвольной ортогональной таблицы $OA(s,n)$ произвольный элемент $y\in Y$ встречается ровно s раз. Поэтому число строк полученной ортогональной таблицы, содержащих элемент y_1, равно $1+(s-1)n$. Так как всего в таблице s^2 строк, то $1+(s-1)n\leqslant s^2$, откуда и следует требуемое неравенство. □

Теорема 11.2 (**достаточные условия оптимального кода аутентификации**). Пусть (X,Y,K,h) — некоторый код аутентификации, для которого выполнены следующие условия:

(i) $|K|=|Y|^2$;

(ii) табличное представление порядка $|K|\times|X|$ над множеством Y хеш-функции h, в которой строки пронумерованы элементами множества K, а столбцы — элементами множества X, является ортогональной таблицей;

(iii) распределение вероятностей на множестве K равномерно.

Тогда данный код аутентификации является оптимальным и распределение вероятностей на множестве Y равномерно.

Доказательство. Заметим, что из условия (ii) следует, что для любых $x,\widetilde{x}\in X$, $x\neq\widetilde{x}$, $y,\widetilde{y}\in Y$ выполнены равенства:

$$|K(x,y)|=|Y|,\quad |K(\widetilde{x},\widetilde{y})\cap K(x,y)|=1.$$

Поэтому, с учетом условий (i) и (iii), для любых $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ будут выполнены такие равенства:

$$P(K(x,y)) = \frac{|K(x,y)|}{|K|} = \frac{|Y|}{|Y|^2} = \frac{1}{|Y|},$$

$$P(K(\widetilde{x},\widetilde{y}) \mid K(x,y)) = \frac{|K(\widetilde{x},\widetilde{y}) \cap K(x,y)|}{|K(x,y)|} = \frac{1}{|Y|}.$$

Для доказательства оптимальности кода аутентификации осталось применить теорему 11.1.

Покажем вторую часть теоремы. Так как для произвольных $x \in X$, $y \in Y$ выполнено равенство $P_{Y|X}(y|x) = P(K(x,y))$, то, применив формулу полной вероятности, получаем:

$$P_Y(y) = \sum_{x \in X} P_X(x) P_{Y|X}(y|x) = \frac{1}{|Y|} \sum_{x \in X} P_X(x) = \frac{1}{|Y|}. \qquad \square$$

Теорема 11.3. Пусть код аутентификации (X, Y, K, h) является оптимальным. Тогда:

(i) $|K| \geqslant |Y|^2$;

(ii) $|K| = |Y|^2$ тогда и только тогда, когда табличное задание хеш-функции h представляет собой ортогональную таблицу $OA(|Y|, |X|)$ и распределение вероятностей $P(K)$ является равномерным.

Доказательство. Из теоремы 11.1 следует, что для любых элементов $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ выполнено равенство:

$$P(K(x,y) \ \cap \ K(\widetilde{x},\widetilde{y})) = \frac{1}{|Y|^2} > 0.$$

Поэтому для любых двух столбцов табличного представления функции h каждая из пар $(y, \widetilde{y}) \in Y^2$ встречается хотя бы один раз. Поэтому:

$$|K| = \sum_{(y,\widetilde{y}) \in Y^2} |K(x,y) \ \cap \ K(\widetilde{x},\widetilde{y})| \geqslant |Y|^2,$$

что показывает справедливость условия (i).

(ii) Если $|K| = |Y|^2$, то из (i) следует, что для любых элементов $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, $y, \widetilde{y} \in Y$ множество $K(x,y) \cap K(\widetilde{x},\widetilde{y})$ состоит в точности из одного элемента. Это означает, что табличное

задание хеш-функции h представляет собой ортогональную таблицу $OA(|Y|, |X|)$ и при фиксированных $x, \widetilde{x} \in X$, $x \neq \widetilde{x}$, для любого $k \in K$ существует единственная пара $y, \widetilde{y} \in Y$, такая что $K(x, y) \cap K(\widetilde{x}, \widetilde{y}) = \{k\}$. Поэтому:

$$P_K(k) = P(K(x, y) \cap K(\widetilde{x}, \widetilde{y})) = \frac{1}{|Y|^2} = \frac{1}{|K|}.$$

Обратно, если табличное задание хеш-функции h представляет собой ортогональную таблицу $OA(|Y|, |X|)$, то очевидно, что $|K| = |Y|^2$. □

Из теоремы 11.3 следует, что минимально возможное число ключей оптимального кода аутентификации равно $|Y|^2$. Следующая теорема описывает все коды аутентификации с минимально возможным числом ключей.

Теорема 11.4. Пусть для некоторого кода аутентификации (X, K, Y, h) выполнено равенство $|K| = |Y|^2$. Код аутентификации (X, K, Y, h) является оптимальным тогда и только тогда, когда выполнены следующие условия:

(i) табличное представление порядка $|K| \times |X|$ над множеством Y хеш-функции h, в которой строки пронумерованы элементами множества K, а столбцы — элементами множества X, является ортогональной таблицей;

(ii) распределение вероятностей на множестве K равномерно.

Доказательство следует из теорем 11.2 и 11.3.

Пусть (X, K, Y, h) — некоторый код аутентификации, в котором $|X| = n$, $K = \{k_1, \ldots, k_r\}$, с распределением вероятностей $P(K)$ на множестве ключей K и табличным заданием хеш-функции A размера $r \times n$ над множеством Y. При этом строки матрицы A пронумерованы элементами множества K, а столбцы — элементами множества X. Пусть также для некоторого другого ключевого множества $\widetilde{K}$, $|\widetilde{K}| \geqslant |K|$, с распределением вероятностей $P(\widetilde{K})$ найдется такое разбиение на r непустых непересекающиеся подмножеств:

$$\widetilde{K} = K_1 \cup K_2 \cup \ldots \cup K_r, \tag{11.3}$$

для которого выполнены равенства:

$$P_{\widetilde{K}}(K_i) = \sum_{k \in K_i} P_{\widetilde{K}}(k) = P_K(k_i), \quad i = 1, \ldots, r. \tag{11.4}$$

Построим код аутентификации $(X, \widetilde{K}, Y, \widetilde{h})$. Как видно, для данного кода остается задать хеш-функцию $\widetilde{h}$. Зададим ее таблично следующим образом: j-ю строку матрицы A продублируем $|K_j|$ раз, $j = 1, \ldots, r$, и из всех полученных (продублированных) строк составим матрицу B. Матрица B и будет табличным заданием хеш-функции $\widetilde{h}$.

Предложение 11.3. Вероятности успехов имитации и успехов подмены для кодов аутентификации (X, K, Y, h) и $(X, \widetilde{K}, Y, \widetilde{h})$ соответственно равны, в частности, из оптимальности одного кода аутентификации следует оптимальность другого.

Данное предложение показывает, что оптимальные коды можно строить не только для случая, когда $P(K)$ равномерно. Пусть:

$$K = K_1 \cup K_2 \cup \ldots \cup K_{s^2} \tag{11.5}$$

— разбиение множества K на непустые непересекающиеся подмножества с условием, что:

$$P_K(K_i) = \sum_{k \in K_i} P_K(k) = \frac{1}{s^2}, \quad i = 1, \ldots, s^2. \tag{11.6}$$

Пусть также для чисел s и n существует ортогональная таблица $OA(s, n)$ над некоторым множеством $Y = \{y_1, \ldots, y_s\}$. Построим из данной таблицы (как и до предложения 11.3) матрицу B размера $|K| \times n$, которая будет таблично представлять хеш-функцию $h : K \times X \to Y$, где $X = \{x_1, \ldots, x_n\}$ — некоторое множество открытых текстов.

Предложение 11.4. Полученный код аутентификации будет являться оптимальным.

Доказательство следует из предложения 11.3 и теоремы 11.4.

Пример 11.2. Пусть:

$$X = \{x_1, x_2, x_3\}, \quad Y = \{0, 1\}, \quad K = \{k_1, \ldots, k_7\}$$

и распределение вероятностей на множестве K имеет вид

K	k_1	k_2	k_3	k_4	k_5	k_6	k_7
P_K	1/16	3/16	1/20	1/10	1/10	1/4	1/4

В этом случае существует разбиение вида (11.5) с условием (11.6):

$$K_1 = \{k_1, k_2\}, \ K_2 = \{k_3, k_4, k_5\}, \ K_3 = \{k_6\}, \ K_4 = \{k_7\},$$

$$P_K(K_1) = P_K(K_2) = P_K(K_3) = P_K(K_4) = \frac{1}{4}.$$

Составим ортогональную таблицу $OA(2,3)$ над Y, а на ее основе построим матрицу B, которая будет являться табличным представлением хеш-функции h :

$OA(2,3)$:

0	0	0
0	1	1
1	0	1
1	1	0

B :

$K \diagdown X$	x_1	x_2	x_3
k_1	0	0	0
k_2	0	0	0
k_3	0	1	1
k_4	0	1	1
k_5	0	1	1
k_6	1	0	1
k_7	1	1	0

Из предложения 11.4 следует, что полученный код аутентификации является оптимальным, причем:

$$P_{im} = P_{podm} = \frac{1}{2}.$$

11.3. Математическая модель кода аутентификации с неограниченным ключом

Пусть A и B — некоторые конечные множества. Будем считать, что сообщения и коды аутентификации являются словами в алфавитах A и B соответственно, т.е. $X \subset A^*$, $Y \subset B^*$, где A^*

и B^* — множество всех слов конечной длины соответственно в алфавитах A и B.

Перед выработкой кода аутентификации сообщение $x \in X$ предварительно представляется в виде последовательности подслов, называемых *кодвеличинами*:

$$x = \underbrace{\ldots}_{u_1}\underbrace{\ldots}_{u_2}\cdots\underbrace{\ldots}_{u_l}$$

В процессе выработки кода аутентификации кодвеличины заменяются некоторыми из эквивалентов в коде аутентификации, которые будем называть *кодобозначениями*.

Пусть U, V — соответственно конечные множества возможных кодвеличин и кодобозначений.

Пусть также имеются конечное множество ключей K и ключевая хеш-функция:

$$h : K \times U \to V.$$

Процесс выработки свертки для сообщения $x = u_1 \ldots u_l$ на ключе $k_1 \ldots k_l$ заключается в замене каждой кодвеличины u_i на кодобозначение v_i в соответствии с ключом k_i, $i = 1, \ldots, l$.

Определение 11.3. *Опорным кодом* кода аутентификации назовем совокупность:

$$\Delta = (U, K, V, h),$$

для которой выполнено равенство:

$$V = \bigcup_{k \in K} h_k(U).$$

Определение 11.4. *l-й степенью опорного кода* Δ назовем совокупность:

$$\Delta^l = (U^l, K^l, V^l, h^{(l)}),$$

где U^l, K^l, V^l — декартовы степени соответствующих множеств U, K, V; множество $h^{(l)}$ состоит из отображений:

$$h_{\overline{k}} : U^l \to V^l, \quad \overline{k} \in K^l,$$

таких что для любых $\overline{u} = u_1 \dots u_l \in U^l$, $\overline{k} = k_1 \dots k_l \in K^l$ выполнено равенство:

$$h_{\overline{k}}(\overline{u}) = h_{k_1}(u_1) \dots h_{k_l}(u_l) = v_1 \dots v_l \in V^l.$$

Обозначим через Δ_H^l следующую совокупность величин:

$$\Delta_H^l = (U^l, K^l, V^l, h^{(l)}, P(U^l), P(K^l)).$$

Определение 11.5. *Кодом аутентификации с неограниченным ключом* назовем семейство:

$$\Delta_H = (\Delta_H^l,\ l \in \mathbb{N};\ \psi_c),$$

где ψ_c — случайный генератор ключевого потока.

Определение 11.6. Код аутентификации с неограниченным ключом Δ_H называется *оптимальным*, если для любого $l \in \mathbb{N}$ код аутентификации Δ_H^l является оптимальным.

Теорема 11.5 (**достаточные условия оптимального кода аутентификации с неограниченным ключом**)**.** Пусть для кода аутентификации Δ_H выполнены следующие условия:

(i) $|K| = |V|^2$;

(ii) для любых $u_1, u_2 \in U$, $v_1, v_2 \in V$ существует, и притом единственный, ключ $k \in K$, для которого выполнены равенства $h_k(u_1) = v_1$ и $h_k(u_2) = v_2$ (данное условие эквивалентно тому, что табличное задание хеш-функции h представляет собой ортогональную таблицу $OA(|V|, |U|)$);

(iii) распределение вероятностей на множестве K равномерно.

Тогда код аутентификации Δ_H является оптимальным и для любого натурального l распределение вероятностей на множестве V^l равномерно.

Доказательство. Зафиксируем произвольное натуральное число l. Из условия (i) следует равенство $|K^l| = |V^l|^2$.

Пусть $\overline{a}, \overline{b} \in U^l$, $\overline{x}, \overline{y} \in V^l$. Тогда из условия (ii) следует, что существует, и притом единственный, ключевой поток $\overline{k} \in K^l$,

что $h_{\overline{k}}(\overline{a}) = \overline{x}$, $h_{\overline{k}}(\overline{b}) = \overline{y}$. А из условия (iii) следует равномерность распределения на множестве K^l (в силу свойств случайного генератора). Поэтому код аутентификации Δ_H является оптимальным в силу теоремы 11.2 и определения 11.6. $\square$

Теорема 11.6. Пусть для кода аутентификации Δ_H выполнено равенство $|K| = |V|^2$. Тогда Δ_H является оптимальным тогда и только тогда, когда выполнены следующие условия:

(i) для любых $u_1, u_2 \in U$, $v_1, v_2 \in V$ существует, и притом единственный, ключ $k \in K$, что $h_k(u_1) = v_1$, $h_k(u_2) = v_2$;

(ii) распределение вероятностей на множестве K равномерно.

Доказательство следует из теорем 11.4 и 11.5.

Пример 11.3. 1. Приведем пример оптимального кода аутентификации с неограниченным ключом. Пусть

$$U = V = V_n, \ K = V_{n^2}$$

и распределение вероятностей $P(K)$ является равномерным. Пусть также табличное задание хеш-функции h представляет собой ортогональную таблицу $OA(|V|, |U|)$. Тогда из теоремы 11.6 следует, что код аутентификации Δ_H является оптимальным.

Для кода аутентификации Δ_H^l, $l \in \mathbb{N}$, обозначим через P_{im}^l вероятность имитации, а через $P_{podm}^l(s)$ — вероятность подмены в сообщении длины l ровно s элементов множества V. Тогда:

$$P_{im}^l = \frac{1}{n^l}, \quad P_{podm}^l(s) = \frac{1}{n^s},$$

т.е. $P_{im}^l \to 0$ при $l \to \infty$, $P_{podm}^l(s) \to 0$ при $s \to \infty$.

2. Заметим, что легко можно строить коды аутентификации с неограниченным ключом, «близкие» к оптимальным. Пусть:

$$U = V = V_n, \ K = V_{2n}$$

и распределение вероятностей $P(K)$ является равномерным. Пусть также табличное представление хеш-функции h размером $2n \times n$, где строки пронумерованы элементами множества K, а столбцы — элементами множества U, имеет следующий

вид. Каждая из первых n строк данной таблицы состоит из одинаковых элементов, причем i-я строка состоит из элементов со значением i, $i = 0, \ldots, n-1$. Остальные n строк данной таблицы образуют латинский квадрат. Тогда для кода аутентификации Δ_H^l, $l \in \mathbb{N}$, будут выполнены следующие выражения:

$$P_{im}^l = \frac{1}{n^l}, \quad P_{podm}^l(s) \leqslant \frac{1}{2^s}.$$

При этом $P_{im}^l \to 0$ при $l \to \infty$, $P_{podm}^l(s) \to 0$ при $s \to \infty$.

Пусть Δ_H — некоторый код аутентификации с неограниченным ключом с опорным кодом $\Delta_H^0 = (U, K, V, h)$, $|U| = n$, $|V| = s$, $|K| = r$, распределением вероятностей $P(K)$ для случайного генератора ψ_c и табличным заданием хеш-функции A размера $r \times n$ над множеством V для кода Δ_H^0. Пусть также для некоторого ключевого множества $\widetilde{K}$, $|\widetilde{K}| = \widetilde{r}$, $\widetilde{r} \geqslant r$, имеется случайный генератор $\widetilde{\psi}_c$ с распределением вероятностей $P_{\widetilde{K}}$ и условием, что найдется разбиение множества $\widetilde{K}$ на r частей вида (11.3) с условием (11.4). Построим код аутентификации с неограниченным ключом $\widetilde{\Delta}_H$ со случайным генератором $\widetilde{\psi}_c$ и опорным кодом $\widetilde{\Delta}_H^0 = (U, \widetilde{K}, V, \widetilde{h})$ со значениями U и V, как и в опорном коде Δ_H^0. Определим хеш-функцию $\widetilde{h}$ с помощью табличного представления B размера $\widetilde{r} \times n$ над множеством V, в которой строки пронумерованы элементами множества $\widetilde{K}$, а столбцы — элементами множества U, следующим образом: j-ю строку матрицы A продублируем $|K_j|$ раз, $j = 1, \ldots, r$, и из всех полученных (продублированных) строк составим матрицу B, которая и будет представлять хеш-функцию $\widetilde{h}$.

Предложение 11.5. Вероятности успехов имитации и успехов подмены для кодов аутентификации Δ_H и $\widetilde{\Delta}_H$ соответственно равны, в частности, из оптимальности одного кода аутентификации следует оптимальность другого.

Пусть имеется разбиение (11.5) с условием (11.6). Пусть также для чисел s и n существует ортогональная таблица $OA(s, n)$ над множеством $V = \{v_1, \ldots, v_s\}$. Построим из данной таблицы

матрицу B размера $s^2 \times n$, которая будет таблично представлять хеш-функцию $h : K \times U \to V$.

Предложение 11.6. Полученный код аутентификации Δ_H будет являться оптимальным.

Глава 12. Электронная подпись

12.1. Общие понятия

Электронная подпись для сообщения — это число, зависящее от самого сообщения и секретного ключа, известного только подписывающему субъекту. Сформулируем основные свойства, которым (в идеале) должна удовлетворять электронная подпись:

- подпись должна быть легко проверяемой и для проверки подписи не требуется знания секретного ключа;
- подписать документ может только «законный» владелец подписи (следовательно, подпись никто не может подделать);
- автор подписи не может от нее отказаться;
- в случае возникновения спора возможно участие третьих лиц (например, суда) для установления подлинности подписи.

Электронная подпись позволяет решить следующие задачи:

- осуществить аутентификацию источника сообщения (получение подтверждения того, что рассматриваемый документ был создан именно указанным источником информации);
- установить целостность сообщения;
- обеспечить невозможность отказа от факта подписи конкретного сообщения.

Электронная подпись имеет много общего с рукописной подписью на бумажном документе. В то же время между обычной и

электронной подписями имеются существенные различия. Сведет эти различия в таблицу.

Рукописная подпись	**Электронная подпись**
Не зависит от подписываемого документа, всегда одинакова	Зависит от подписываемого документа, практически всегда разная
Связана с подписывающим субъектом, не может быть утеряна	Определяется секретным ключом, принадлежащим подписывающему лицу, может быть утеряна владельцем
Неотделима от носителя (бумаги), поэтому подписывается каждый экземпляр документа	Легко отделима от документа, поэтому верна для всех его копий
Не требует для реализации дополнительных механизмов	Требует дополнительных механизмов, реализующих алгоритмы ее вычисления и проверки
Не требует создания поддерживающей инфраструктуры	Требует создания доверенной инфраструктуры сертификатов открытых ключей

Для проверки электронной подписи необходима некоторая открытая информация (открытые ключи) для обеспечения возможности открытой проверки. Для исключения возможности подделки этой информации лицами, которые хотят выступать от лица законного владельца подписи (секретного ключа), создается *инфраструктура открытых ключей*, состоящая из центров сертификации открытых ключей и обеспечивающая своевременного подтверждения достоверности принадлежности данной открытой информации заявленному владельцу и обнаружения подлога.

12.2. Электронная подпись RSA

Данная электронная подпись базируется на асимметричном шифре RSA. Каждый, кто хочет подписывать свои сообщения подписью RSA, должен выбрать два больших простых числа p и q, вычислить $n = pq$, $\varphi(n) = (p-1)(q-1)$ — значение функции Эйлера от n, зафиксировать некоторое число $e < \varphi(n)$, взаимно простое с $\varphi(n)$, и найти решение сравнения $ed \equiv 1 \pmod{\varphi(n)}$ относительно неизвестного значения d (напомним, что данное сравнение имеет единственное решение). После этого параметры e и n каждого абонента размещаются в открытом справочнике, а значения d, p и q держатся в секрете.

Пусть абонент A хочет подписать сообщение m. Алгоритм вычисления электронной подписи RSA имеет следующий вид:

1. Абонент A вычисляет значение хеш-функции $h = h(m)$, алгоритм которой должен быть общедоступным.
2. A вычисляет значение $s \equiv h^d \pmod{n}$.
3. Подписанное сообщение имеет вид (m, s).

Для того, чтобы проверить подлинность подписи, требуется проделать следующие шаги:

1. Вычисляется значение $h = h(m)$.
2. Проверяется справедливость сравнения $h \equiv s^e \pmod{n}$; если оно верно, то подпись принимается; если — нет, то отвергается.

Пример 12.1 (банки и вкладчики)**.** Рассмотрим задачу, в которой вкладчики банка v_1, v_2, ..., v_s передают шифрованное распоряжение работнику банка B (банкиру). При этом кроме конфиденциальности должна обеспечиваться узнаваемость вкладчика, чтобы по полученному сообщению банкир B сумел идентифицировать автора сообщения и выполнить именно его распоряжение.

Приведем решение данной задачи на основе шифра RSA. Банкир B выбирает некоторые большие простые числа P, Q и вычисляет $N = PQ$. Каждый из вкладчиков v_i, $i = 1, \ldots, s$ также выбирает свои значения p_i, q_i, $n_i = p_i q_i$, причем желательно, чтобы $N > n_i$. Затем как банкир B, так и вкладчики находят значения $\varphi(N) = (P-1)(Q-1)$ и $\varphi(n_i) = (p_i - 1)(q_i - 1)$. После этого каждый выбирает свой открытый ключ: E — банкир, e_i — вкладчики из условий:

$$1 < E < \varphi(N), \quad (E, \varphi(N)) = 1,$$

$$1 < e_i < \varphi(n_i), \quad (e_i, \varphi(n_i)) = 1.$$

Затем банкир и вкладчики находят свои секретные ключи D и d_i из сравнений:

$$ED \equiv 1 \pmod{\varphi(N)}, \quad 1 < D < \varphi(N),$$

$$e_i d_i \equiv 1 \pmod{\varphi(n_i)}, \quad 1 < d_i < \varphi(n_i).$$

После этих операций открыто публикуется справочник с открытыми ключами:

$$B: \ N, \ E,$$

$$v_i: \ n_i, \ e_i.$$

Пусть некоторый вкладчик v_i хочет передать распоряжение m, $m < \min\{n_i, N\}$, банкиру B. Он шифрует (подписывает) его сначала своим секретным ключом:

$$m_1 \equiv m^{d_i} \pmod{n_i},$$

а затем открытым ключом банкира B :

$$m_2 \equiv m_1^E \pmod{N}.$$

Сообщение m_2 передается по открытому каналу связи. Банкир, получив сообщение m_2, сначала расшифровывает его своим секретным ключом D :

$$m_3 \equiv m_2^D \pmod{N},$$

а затем открытым ключом вкладчика v_i :

$$m_4 \equiv m_3^{e_i} \pmod{n_i}.$$

При этом $m_4 \equiv m \pmod{n_i}$.

Если вкладчик из открытого справочника узнает, что $n_i > N$, то, изменив порядок шифровки, получит тот же результат, если банкир также изменит порядок шифровки.

12.3. Электронная подпись Фиата-Шамира

Данная электронная подпись основана на сложности задачи факторизации больших чисел и извлечения квадратного корня в кольце целых чисел.

Пусть A — алфавит, в котором записываются сообщения и которые требуется подписывать, $X \subseteq A^*$, $h : X \to V_m$ — некоторая хеш-функция, алгоритм которой не держится в секрете. Значением данной хеш-функции для любого сообщения из X является последовательность из m бит.

Чтобы подписывать сообщения подписью Фиата-Шамира, необходимо выбрать достаточно большие простые числа p и q, вычислить $n = pq$ и сгенерировать некоторый набор из m чисел $a_1, \ldots, a_m \in \mathbb{Z}_n$ с условием, что для любого $i = 1, \ldots, m$ числа a_i и n должны быть взаимно простыми (это гарантирует существование $a_i^{-1} \in \mathbb{Z}_n$). Набор чисел $a_1, \ldots, a_m$ является секретным ключом. На основе данного секретного ключа вычисляется открытый ключ, в качестве которого выступает число n и набор из m чисел кольца вычетов $\mathbb{Z}_n$:

$$b_i = (a_i^{-1})^2 \pmod{n}, \quad i = 1, \ldots, n.$$

Каждый участник публикует в открытом справочнике свой открытый ключ n, b_1, $\ldots$,b_m.

Предположим, что требуется подписать некоторое сообщение $M \in X$. Алгоритм вычисления электронной подписи Фиата-Шамира имеет следующий вид:

1. Выбирается случайное целое число k $(0 < k < n)$ и вычисляется значение $r = k^2 \pmod{n}$.

2. Вычисляется значение хеш-функции:

$$h = h(M, r) = h_1 \ldots h_m \in V_m.$$

3. Вычисляется значение $s = ka_1^{h_1} \dots a_m^{h_m} \pmod n$.

4. Подписанное сообщение имеет вид (M, h, s).

Для того, чтобы проверить подлинность подписи, требуется проделать следующие шаги:

1. Найти в справочнике отрытый ключ n, b_1, $\dots$,b_m и вычислить значение:
$$\widetilde{r} = s^2 b_1^{h_1} \dots b_m^{h_m} \pmod n.$$

2. Вычислить значение хеш-функции $\widetilde{h} = h(M, \widetilde{r})$.

3. Проверяется равенство $h = \widetilde{h}$; если оно верно, то подпись принимается; если — нет, то отвергается.

Справедливость такой проверки следует из равенств:

$$\widetilde{r} = s^2 b_1^{h_1} \dots b_m^{h_m} =$$
$$= (ka_1^{h_1} \dots a_m^{h_m})^2 (a_1^{-1})^{2h_1} \dots (a_m^{-1})^{2h_m} \equiv k^2 \pmod n.$$

12.4. Электронная подпись Эль-Гамаля

Опишем электронную подпись на базе шифра Эль-Гамаля, основанную на сложности задачи дискретного логарифмирования. Каждый, кто хочет подписывать свои сообщения подписью Эль-Гамаля, должен выбрать некоторое простое число p и некоторый первообразный корень g по модулю p. Также выбирается достаточно большое число x, $1 < x < p - 1$, которое держится в секрете. С помощью секретного ключа вычисляется значение открытого ключа: $y = g^x \pmod p$. Числа p, g и открытый ключ y размещаются в открытом справочнике.

Предположим, что требуется подписать некоторое сообщение M. Алгоритм вычисления электронной подписи Эль-Гамаля имеет следующий вид:

1. Вычисляется значение хеш-функции $h = h(M)$, алгоритм которой должен быть общеизвестным, причем для значения свертки h должно выполняться неравенство $1 < h < p$.

2. Генерируется случайным образом некоторое число k, $1 < k < p-1$, взаимно простое с числом $p-1$ (заметим, что требование $(k, p-1) = 1$ не требовалось в шифре Эль-Гамаля), и вычисляются значения:

$$r \equiv g^k \pmod{p},$$

$$s \equiv k^{-1}(h - xr) \pmod{p-1}.$$

3. Подписанное сообщение имеет вид (M, r, s).

Для того, чтобы проверить подлинность подписи, требуется проделать следующие шаги:

1. Вычисляется значение хеш-функции $h = h(M)$.

2. Проверяется справедливость сравнения:

$$y^r r^s \equiv g^h \pmod{p};$$

если оно верно, то подпись принимается; если — нет, то отвергается.

Справедливость такой проверки следует из соотношений:

$$y^r r^s = g^{xr} g^{kk^{-1}(h-xr)} \equiv g^h \pmod{p}.$$

12.5. Электронная подпись Шнорра

Схема электронной подписи Шнорра основана на сложности вычисления значения логарифма в конечном поле.

Пусть p — некоторое простое число, q — простой делитель числа $p-1$, $g \in \mathbb{Z}_p$, имеющий порядок q, h — хеш-функция. Выберем случайное число x в интервале $1 \leqslant x \leqslant q-2$ и вычислим значение $y = g^{-x} \pmod{p}$.

Число x является секретным ключом, набор p, q, g, y — открытым ключом.

Опишем алгоритм вычисления электронной подписи для сообщения M :

1. Генерируется случайное целое число k $(1 \leqslant k \leqslant q-2)$ и вычисляется $r = g^k \pmod{p}$.

2. Вычисляется свертка $h = h(M, r)$.

3. Вычисляется $s = k + xh \pmod{q}$.

4. Подписанное сообщение имеет вид (M, h, s).

Рассмотрим теперь алгоритм проверки подписи:

1. Вычисляется $\widetilde{r} = g^s y^h \pmod{p}$.

2. Вычисляется $\widetilde{h} = h(M, \widetilde{r})$.

3. Проверяется равенство $h = \widetilde{h}$; если оно верно, то подпись принимается; если — нет, то отвергается.

12.6. Электронная подпись на основе эллиптических кривых

12.6.1. Электронная подпись ГОСТ Р 34.10-2012

Российский стандарт электронной подписи ГОСТ Р 34.10-2012 [15] позволяет работать как с ключами электронной подписи длины 256 бит (длина открытого ключа — 512 бит), так и с ключами длины 512 бит (длина открытого ключа — 1024 бита) и должен использоваться исключительно совместно с алгоритмом хеширования ГОСТ Р 34.11-2012 (см. параграф 10.5).

Российский стандарт электронной подписи ГОСТ Р 34.10-2012, как и его предшественник ГОСТ Р 34.10-2001, основан на вычислениях в группе точек эллиптических кривых. Можно сказать, что ГОСТ Р 34.10-2001 и ГОСТ Р 34.10-2012 являются аналогами ГОСТ Р 34.10-94 при замене класса используемых циклических групп с мультипликативных групп простых полей на группы точек эллиптических кривых. Схема ГОСТ Р 34.10-94 по структуре преобразований, в свою очередь, близка к классической схеме подписи Эль-Гамаля. Стойкость электронной подписи ГОСТ Р 34.10-2012 основывается на сложности вычисления дискретного логарифма в группе точек эллиптической кривой, а также на стойкости используемой хеш–функции ГОСТ Р 34.11-2012. ГОСТ Р 34.10-2012 должен использоваться исключительно совместно с алгоритмом хеширования ГОСТ Р

34.11-2012 (см. параграф 10.5). Электронная подпись представляется в виде двоичного вектора длиной 512 или 1024 бита.

В данном стандарте используются эллиптические кривые $E_p(a, b)$ вида:

$$y^2 = x^3 + ax + b \pmod p, \quad a, b \in \mathbb{Z}_p,$$

над полем $\mathbb{Z}_p$, $p > 3$, где $4a^3 + 27b^2 \not\equiv 0 \pmod p$. Напомним, что относительно операции сложения (определенной в параграфе 3.3) множество всех точек эллиптической кривой $E_p(a, b)$, вместе с нулевой точкой, образуют конечную аддитивную абелеву группу порядка n, для которого выполнено неравенство Хассе:

$$p + 1 - 2\sqrt{p} \leqslant n \leqslant p + 1 + 2\sqrt{p}.$$

Особенностью стандарта ГОСТ Р 34.10-2012 является то, что в документе не зафиксированы какие-либо кривые, рекомендуемые для использования, присутствует только набор требований к ним.

Напомним, что *инвариантом* эллиптической кривой $E = E_p(a, b)$ называется величина $j(E)$, удовлетворяющая сравнению:

$$j(E) \equiv 1728\frac{4a^3}{4a^3 + 27b^2} \pmod p.$$

Из этого равенства видно, что кривые с коэффициентом $a = 0$ — кривые с нулевым j-инвариантом, а кривые с коэффициентом $b = 0$ — кривые с j-инвариантом, равным 1728.

Помимо того, что эллиптическую кривую $E_p(a, b)$ можно задать с помощью определения параметров a и b, можно ее определить по известному инварианту следующим образом:

$$\begin{cases} a \equiv 3k \pmod p, \\ b \equiv 2k \pmod p, \end{cases}$$

где:

$$k \equiv \frac{j(E)}{1728 - j(E)} \pmod p, \quad j(E) \neq 0,\ j(E) \neq 1728.$$

Параметры электронной подписи:

1. Простое число p — модуль эллиптической кривой.
2. Эллиптическая кривая $E_p(a, b)$, задаваемая коэффициентами $a, b \in \mathbb{Z}_p$ или инвариантом $j(E)$.
3. Целое число n — порядок группы точек эллиптической кривой $E_p(a, b)$.
4. Простое число q — порядок циклической группы (образованной некоторой точкой $G \in E_p(a, b)$) точек эллиптической кривой $E_p(a, b)$, для которого выполнены следующие условия: $q|n$ и $2^{254} < q < 2^{256}$ (если длина свертки хеш-функции равна 256) или $2^{508} < q < 2^{512}$ (если длина свертки равна 512).
5. Точка $G = G(x_G, y_G) \in E_p(a, b)$ порядка q.
6. Хеш-функция $h : V^* \to V_m$ (отображающая сообщения, представленные в виде двоичных векторов произвольной конечной длины, в двоичные вектора длины m бит), определенная стандартом ГОСТ Р 34.11-2012. Если $m = 256$, то $2^{254} < q < 2^{256}$. Если $m = 512$, то $2^{508} < q < 2^{512}$.

Пусть x — некоторое целое число, $0 < x < q$, $Y = [x]G = (x_Y, y_Y) \in E_p(a, b)$. Число x является секретным ключом подписи, Y — открытый ключ для проверки подписи.

К приведенным выше параметрам схемы электронной подписи предъявляют следующие требования:

- должно быть выполнено условие $p^t \not\equiv 1 \pmod{q}$ для всех целых $t = 1, 2, \ldots, B$, где $B = 31$, если $2^{254} < q < 2^{256}$, и $B = 131$, если $2^{508} < q < 2^{512}$;
- должно быть выполнено неравенство $n \neq p$;
- инвариант кривой должен удовлетворять условию $j(E) \neq 0$, $j(E) \neq 1728$.

Формирование подписи. Для получения электронной подписи под сообщением $M \in V^*$ необходимо выполнить следующие действия.

1. Вычислить свертку сообщения M : $h = h(M) \pmod q$. Будем через h обозначать и свертку сообщения — двоичный вектор, и двоичное представление соответствующего числа. Если $h = 0$, то определить $h = 1$.

2. Сгенерировать случайное (псевдослучайное) целое число k, удовлетворяющее неравенству $0 < k < q$.

3. Вычислить точку эллиптической кривой:

$$R = [k]G = (x_R, y_R) \in E_p(a, b)$$

и определить значение $r = x_R \pmod q$. Если $r = 0$, то вернуться к шагу 2.

4. Вычислить значение $s = (rx + kh) \pmod q$. Если $s = 0$, то вернуться к шагу 2.

5. Подписанное сообщение имеет вид (M, r, s).

Проверка электронной подписи. Для проверки электронной подписи необходимо проделать следующие действия:

1. Проверить неравенства $0 < r < q$, $0 < s < q$. Если хотя бы одно неверно, то подпись неверна.

2. Вычислить свертку сообщения M : $h = h(M) \pmod q$. Если $h = 0$, то определить $h = 1$.

3. Вычислить значение $h^{-1} \pmod q$ и определить:

$$z_1 = sh^{-1} \pmod q, \quad z_2 = -rh^{-1} \pmod q.$$

4. Вычислить точку эллиптической кривой $\widetilde{R} = [z_1]G + [z_2]Y = (x_{\widetilde{R}}, y_{\widetilde{R}}) \in E_p(a, b)$ и определить $\widetilde{r} = x_{\widetilde{R}} \pmod q$.

5. Если выполнено равенство $r = \widetilde{r}$, то подпись принимается, в противном случае — подпись неверна.

Справедливость такой проверки следует из соотношений:

$$\widetilde{R} = [sh^{-1} \pmod q]G + [-rh^{-1} \pmod q]Y =$$
$$= [sh^{-1} \pmod q]G + [-rh^{-1}x \pmod q]G =$$
$$= [h^{-1}(s - rx) \pmod q]G = [k]G = R.$$

12.6.2. Электронная подпись ECDSA

Алгоритм ECDSA в 1999 г. был принят как стандарт ANSI, в 2000 г. — как стандарт IEEE и NIST. Также в 1998 г. алгоритм был принят стандартом ISO.

Параметры электронной подписи:

1. Простое число p — модуль эллиптической кривой.
2. Эллиптическая кривая $E_p(a, b)$, задаваемая коэффициентами $a, b \in \mathbb{Z}_p$, причем $4a^3 + 27b^2 \not\equiv 0 \pmod{p}$.
3. Целое число n — порядок группы точек эллиптической кривой $E_p(a, b)$. n должно делиться на достаточно большое простое число q. Стандарт ANSI X9.62 требует, чтобы $q > 2^{160}$ и $q > 4\sqrt{p}$.
4. Простое число q — порядок циклической группы (образованной некоторой точкой $G \in E_p(a, b)$) точек эллиптической кривой $E_p(a, b)$, для которого выполнены следующие условия: $q|n$, $q > 2^{160}$ и $q > 4\sqrt{p}$. Согласно ANSI X9.62 q не должно быть делителем числа $p^k - 1$ для любого $k = 1, \ldots, 100$. Также должно выполняться условие $q \neq p$.
5. Точка $G = G(x_G, y_G) \in E_p(a, b)$ порядка q.
6. Хеш-функция $h : V^* \to V_m$ (отображающая сообщения, представленные в виде двоичных векторов произвольной конечной длины, в двоичные вектора длины m бит).

Пусть x — некоторое целое число, $0 < x < q$, $Y = [x]G = (x_Y, y_Y) \in E_p(a, b)$. Число x является секретным ключом подписи, Y — открытый ключ для проверки подписи.

Формирование подписи. Для получения электронной подписи под сообщением $M \in V^*$ необходимо выполнить следующие действия:

1. Вычислить свертку сообщения $M : h = h(M)$. Если $q < h$, то используются только левые биты значения хеш-функции.

2. Сгенерировать случайное (псевдослучайное) целое число k, удовлетворяющее неравенству $0 < k < q$.

3. Вычислить точку эллиптической кривой

$$R = [k]G = (x_R, y_R) \in E_p(a, b)$$

и определить значение $r = x_R \pmod q$. Если $r = 0$, то вернуться к шагу 2.

4. Вычислить значение $k^{-1} \pmod q$ и определить $s = k^{-1}(h + rx) \pmod q$. Если $s = 0$, то вернуться к шагу 2.

5. Подписанное сообщение имеет вид (M, r, s).

Проверка электронной подписи. Для проверки электронной подписи необходимо проделать следующие действия:

1. Проверить неравенства $0 < r < q$, $0 < s < q$. Если хотя бы одно неверно, то подпись неверна.

2. Вычислить свертку сообщения $M : h = h(M)$. Если $q < h$, то используются только левые биты значения хеш-функции.

3. Вычислить значения $z_1 = s^{-1}h \pmod q$, $z_2 = s^{-1}r \pmod q$.

4. Вычислить точку эллиптической кривой $\widetilde{R} = [z_1]G + [z_2]Y = (x_{\widetilde{R}}, y_{\widetilde{R}}) \in E_p(a, b)$ и определить $\widetilde{r} = x_{\widetilde{R}} \pmod q$.

5. Если выполнено равенство $r = \widetilde{r}$, то подпись принимается, в противном случае — подпись неверна.

Справедливость такой проверки следует из соотношений:

$$\widetilde{R} = [s^{-1}h \pmod q]G + [s^{-1}r \pmod q]Y =$$

$$= [s^{-1}h \pmod q]G + [s^{-1}rx \pmod q]Y =$$

$$= [k(h + rx)^{-1}(h + rx) \pmod q]G = [k]G = R.$$

12.7. Электронная подпись Диффи-Лампорта на основе симметричных систем шифрования

Пусть требуется подписать двоичное сообщение $m = m_1 \dots m_n$, $m_i \in \{0,1\}$, $i = 1, \dots, n$. Подписывающий сначала выбирает n пар случайных секретных ключей:

$$K = \{(k_{10}, k_{11}), \dots, (k_{n0}, k_{n1})\}$$

для симметричного шифрования E_k и n пар случайных чисел:

$$S = \{(s_{10}, s_{11}), \dots, (s_{n0}, s_{n1})\}, \quad s_{ij} \in \{0,1\},$$

$$i = 1, \dots, n, \quad j = 0, 1,$$

и вычисляет значения:

$$r_{ij} = E_{k_{ij}}(s_{ij}), \quad i = 1, \dots, n, \quad j = 0, 1.$$

Наборы S и $R = \{(r_{10}, r_{11}), \dots, (r_{n0}, r_{n1})\}$ являются открытыми и размещаются в общедоступном справочнике.

Подпись для сообщения m имеет вид $(k_{1m_1}, \dots, k_{nm_n})$. Для проверки электронной подписи следует проверить равенства:

$$r_{ij} = E_{k_{ij}}(s_{ij}), \quad j = m_i, \quad i = 1, \dots, n.$$

Так как после вычисления одной подписи часть секретного ключа становится известной, данная электронная подпись является одноразовой.

Еще одним недостатком такой схемы является слишком большой размер подписи. Чтобы устранить этот недостаток, для этого имеется несколько способов:

1. Можно подписывать не само сообщение, а его свертку, которая будет иметь всегда фиксированную длину. Для этого нужно воспользоваться какой-либо хеш-функцией.

2. В компьютере подписывающего можно хранить не $2n$ значений секретных ключей, а лишь один секретный (первичный) ключ k, а (вторичные) ключи k_{ij}, с помощью которых получаются значения r_{ij}, вычислять по следующему правилу:

$$k_{ij} = E_k(i, j), \quad i = 1, \dots, n, \quad j = 0, 1.$$

3. Набор открытых значений S также можно свернуть, например, s_{ij} можно получить как сумму битов двоичного вектора $E_k(i, j)$ по модулю два.

Глава 13. Схемы разделения секрета

13.1. Пороговые схемы разделения секрета

Пусть имеются n участников $A_1,\ldots,A_n$, между которыми требуется разделить некоторый секрет S, и некоторый выделенный участник D, называемый *дилером*, который разделяет секрет S. Пусть также t — некоторое натуральное число, причем $1 \leqslant t \leqslant n$, $\{\alpha_1,\ldots,\alpha_n\}$ — некоторая информация о секрете S, и выполнены следующие условия:

1. Каждый участник A_i знает некоторую информацию α_i, которая неизвестна остальным $n-1$ участникам.
2. Секрет S легко может быть вычислен по произвольному t-элементному подмножеству множества $\{\alpha_1,\ldots,\alpha_n\}$.
3. Секрет S нельзя вычислить ни по какому $(t-1)$-элементному подмножеству в $\{\alpha_1,\ldots,\alpha_n\}$.

В этом случае множество $\{\alpha_1,\ldots,\alpha_n\}$ со свойствами 1–3 называется (n,t) *пороговой схемой* для секрета S.

Если $n=t$, то задачу разделения секрета можно решить следующим образом. Пусть G — некоторая аддитивная абелева группа (например, $\mathbb{Z}$, $\mathbb{Z}_p$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ и т.д.) и требуется между n участниками разделить секрет $S = s_1 s_2 \ldots s_k$, где все $s_i \in G$. Сгенерируем $n-1$ случайную последовательность элементов группы G, каждая из которых имеет длину k :

$$\begin{array}{cccc} a_{11} & a_{12} & \ldots & a_{1k}, \\ a_{21} & a_{22} & \ldots & a_{2k}, \\ & & \ldots & \\ a_{n-1,1} & a_{n-1,2} & \ldots & a_{n-1,k}. \end{array}$$

Каждую из данных последовательностей передадим $n-1$ участнику, а оставшемуся участнику передадим такую последовательность:

$$a_{n1} = s_1 - a_{11} - a_{21} - \ldots - a_{n-1,1},$$

$$a_{n2} = s_2 - a_{12} - a_{22} - \ldots - a_{n-1,2},$$

$$\ldots$$

$$a_{nk} = s_k - a_{1k} - a_{2k} - \ldots - a_{n-1,k}.$$

В этом случае секрет $S = s_1 \ldots s_k$ можно получить, если все n участников предоставят свои части секрета:

$$s_1 = a_{11} + a_{21} + \ldots + a_{n1},$$

$$\ldots$$

$$s_k = a_{1k} + a_{2k} + \ldots + a_{nk}.$$

При этом заметим, что «+» и «−» — бинарные операции аддитивной абелевой группы G.

13.1.1. Схема разделения секрета Шамира

Пусть F — некоторое поле. Данная схема основана на следующем хорошо известном из курса линейной алгебры утверждении.

Предложение 13.1. Пусть имеется $n+1$ пара элементов (x_0, y_0), (x_1, y_1),. . . ,(x_n, y_n), где $x_i, y_i \in F$, $i = 0, 1, \ldots, n$, причем $x_i \neq x_j$ при $i \neq j$. Тогда существует, и притом единственный, многочлен $L(x)$ степени не более n, для которого $L(x_i) = y_i$, $i = 0, 1, \ldots, n$.

Многочлен $L(x)$ носит название *многочлена Лагранжа*, который можно записать в следующем виде:

$$L(x) = \sum_{i=0}^{n} y_i \prod_{\substack{0 \leqslant j \leqslant n \\ j \neq i}} \frac{x - x_j}{x_i - x_j}. \tag{13.1}$$

Схема разделения секрета Шамира позволяет создать (n, t) пороговую схему для любых n и t, где $1 \leqslant t \leqslant n$.

Пусть F — некоторое поле и $S \in F$ — некоторый секрет. Зафиксируем произвольным образом n различных несекретных элементов $x_1, \ldots, x_n \in F$ с условием, что $x_i \neq x_j$ при $i \neq j$. Дилер D выбирает случайным образом $t-1$ секретных коэффициентов $a_1, \ldots, a_{t-1} \in F$, причем $a_{t-1} \neq 0$, многочлена:

$$P_{t-1}(x) = S + a_1 x + a_2 x^2 + \ldots + a_{t-1} x^{t-1}. \qquad (13.2)$$

После этого дилер вычисляет значения:

$$y_i = P_{t-1}(x_i), \quad i = 1, \ldots, n.$$

Каждому участнику A_i дилер передает пары (x_i, y_i), $i = 1, \ldots, n$.

В этом случае из предложения 13.1 следует, что многочлен $P_{t-1}(x)$ можно однозначно восстановить по произвольному t-элементному подмножеству множества $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ и нельзя этого сделать ни для какого $(t-1)$-элементного подмножества. При этом для восстановления $P_{t-1}(x)$ можно воспользоваться формулой (13.1). Таким образом, рассматриваемое множество $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ является (n, t) пороговой схемой для секрета S.

Заметим, что для любого t-элементного подмножества:

$$\{(\widetilde{x}_1, \widetilde{y}_1), \ldots, (\widetilde{x}_t, \widetilde{y}_t)\}$$

в $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ из формулы (13.1) следует такое равенство:

$$S = P_{t-1}(0) = \sum_{i=1}^{t} \widetilde{y}_i \prod_{\substack{1 \leqslant j \leqslant t \\ j \neq i}} \frac{\widetilde{x}_j}{\widetilde{x}_j - \widetilde{x}_i}.$$

При этом очень часто (если позволяет структура поля F) в качестве значений x_i берут такие значения: $x_i = i \in F$, $i = 1, \ldots, n$.

Схема Шамира обладает следующими основными свойствами:

- Совершенная секретность (теоретико-информационная стойкость): наличие любых m долей секрета, $m < t$, не дает никакой информации о секрете.

- Идеальность: число битов, содержащихся в каждой доле секрета, равно числу битов, содержащихся в самом секрете.
- Расширяемость: число владельцев долей секрета n может быть в любой момент увеличено вплоть до значения $|F|-1$. При этом количество частей секрета t, необходимых для восстановления секрета, останется неизменным.
- Гибкость: возможно присваивать различные «веса» различным подмножествам авторизации. Например, для $(n,3)$-пороговой схемы директор получает все три доли, каждый замдиректора получает две и, наконец, каждый простой участник схемы разделения получает одну долю.
- Свойство гомоморфизма. Пусть S и R — секреты. Данные секреты разделены на части: $x_1, P_{t-1}(x_1),\ldots,\ x_n, P_{t-1}(x_n)$ и $x_1, Q_{t-1}(x_1),\ldots,\ x_n, Q_{t-1}(x_n)$, $S = P_{t-1}(0)$, $R = Q_{t-1}(0)$. Если каждый участник просуммирует $H_{t-1}(x_i) = P_{t-1}(x_i) + Q_{t-1}(x_i)$, $i = 1,\ldots,n$, то каждая из полученных сумм, в свою очередь, является частью секрета $S+R$, определяемого из полинома $H_{t-1}(x) = P_{t-1}(x) + Q_{t-1}(x)$, причем $H_{t-1}(0) = S+R$.

Замечание 13.1. Предположим, что в роли секрета выступает некоторый (произвольный) файл (текстовый, графический и т.д.). Хорошо известно, что любой файл рассматривается как последовательность байт, оканчивающаяся маркером конца файла. В этом случае файл разбивается на последовательность $S_1,\ S_2,\ldots,\ S_l,\ S_i \in \mathbb{Z}_{2^{8k}}$ для некоторого фиксированного натурального k, $i = 1,\ldots,l$. Например, при $k=1$ файл разбивается на последовательность байт, при $k=2$ — на последовательность блоков по два байта и т.д. Очень часто при использовании схемы разделения секрета Шамира в этом случае выбирается простое число $p > \max\{2^{8k}, n\}$ и при разделении секретов S_i на части все операции проходят в поле $\mathbb{Z}_p$ (кольцо вычетов по простому модулю p). Поэтому при разделении секретов S_i на n участников при заданном пороге t будут получаться доли секрета, принадлежащие полю $\mathbb{Z}_p$. Так как $p > 2^{8k}$, то для долей секрета

S_i понадобится больший объем памяти, чем для самого секрета S_i. Это противоречит свойству идеальности схемы Шамира. Выходом из данной ситуации является использование конечных полей (полей Галуа) вида $GF(2^{8k})$. В этом случае свойство идеальности будет выполнено.

13.1.2. Проверяемая схема Фельдмана-Шамира и ее модификация на эллиптических кривых

Рассмотрим ситуацию, когда участники разделения секрета D, $A_1,\ldots,$ A_n не доверяют друг другу. Пусть p, q — большие простые числа, причем q — делитель числа $p-1$, g — некоторый элемент мультипликативной группы $\mathbb{Z}_p^*$, имеющий порядок q. При этом заметим, что в качестве поля F в схеме Шамира будет выступать поле $\mathbb{Z}_q$. После построения дилером D многочлена (13.2) над полем $\mathbb{Z}_q$ он вычисляет значения:

$$r_i = g^{a_i} \pmod p, \quad i = 0, 1, \ldots, t-1,$$

где $a_0 = S$. После этого значения r_i, $i = 0, 1, \ldots, t-1$, размещаются в открытом доступе. Сложность вычисления значений a_i по известным r_i основана на сложности задачи дискретного логарифмирования в конечном поле.

После получения участником A_i пары (x_i, y_i) (доли секрета S) он может проверить, что полученная от дилера информация (x_i, y_i) действительно является долей секрета S, проверив сравнение:

$$g^{y_i} \equiv r_0 \left(r_1\right)^{x_i} \ldots \left(r_{t-1}\right)^{x_i^{t-1}} \pmod p,$$

так как:

$$g^{y_i} = g^{P_{t-1}(x_i)} = g^S g^{a_1 x_i} \ldots g^{a_{t-1} x_i^{t-1}} \equiv$$
$$\equiv r_0 \left(r_1\right)^{x_i} \ldots \left(r_{t-1}\right)^{x_i^{t-1}} \pmod p.$$

Также после восстановления секрета S любой группой из t рассматриваемых участников истинность значения S (например, некоторый участник может предоставить заведомо ложную долю секрета) можно проверить с помощью сравнения:

$$g^S \equiv r_0 \pmod p.$$

Приведем модификацию схемы Фельдмана-Шамира на эллиптических кривых. Пусть q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$, где $E_p(a,b)$ — эллиптическая кривая над полем $\mathbf{Z}_p$ вида:

$$E_p(a,b): \ y^2 = x^3 + ax + b \ (\mathrm{mod}\ p).$$

Пусть некоторая точка $G \in E_p(a,b)$ имеет порядок q, т.е. образует циклическую подгруппу порядка q в $(E_p(a,b),+)$:

$$\langle G\rangle = \{G,\ [2]G, \ldots, [q]G = \mathcal{O}\}.$$

После построения дилером D многочлена (13.2) над полем $\mathbb{Z}_q$ он вычисляет точки эллиптической кривой:

$$R_i = [a_i]G, \quad i = 0, 1, \ldots, t-1,$$

где $a_0 = S$. После этого значения R_i, $i = 0, 1, \ldots, t-1$, размещаются в открытом доступе. Сложность вычисления значений a_i по известным R_i основана на сложности задачи дискретного логарифмирования в группе точек эллиптической кривой.

После получения участником A_i пары (x_i, y_i) (доли секрета S) он может проверить, что полученная от дилера информация (x_i, y_i) действительно является долей секрета S, проверив равенство:

$$[y_i]G = R_0 + [x_i]R_1 + \ldots + [x_i^{t-1}]R_{t-1},$$

так как:

$$[y_i]G = [P_{t-1}(x_i)]G = [S]G + [x_i]([a_1]G) + \ldots + [x_i^{t-1}]([a_{t-1}]G) =$$

$$= R_0 + [x_i]R_1 + \ldots + [x_i^{t-1}]R_{t-1}.$$

Также после восстановления секрета S любой группой из t рассматриваемых участников истинность значения S можно проверить с помощью равенства:

$$[S]G = R_0.$$

13.1.3. Совершенная проверяемая схема Педерсена-Шамира и ее модификация на эллиптических кривых

Если схема Фельдмана-Шамира основана на трудной задаче дискретного логарифмирования, то приводимая ниже проверяемая схема имеет совершенную секретность (теоретико-информационную стойкость).

Пусть числа p, q, g, S определяются также, как и в предыдущей схеме, $d \in \mathbb{Z}_q$ — некоторый секретный параметр, $h = g^d \pmod p$ — открытый параметр.

Для разделения секрета $S \in \mathbb{Z}_q$ дилер D выбирает два многочлена $P_{t-1}(x)$ и $Q_{t-1}(x)$ степени $t-1$ над полем $\mathbb{Z}_q$:

$$P_{t-1}(x) = S + a_1 x + \ldots + a_{t-1}x^{t-1} \in \mathbb{Z}_q[x],$$

$$Q_{t-1}(x) = b_0 + b_1 x + \ldots + b_{t-1}x^{t-1} \in \mathbb{Z}_q[x],$$

где $a_0 = S$, a_i, b_j — случайные числа из $\mathbb{Z}_q$, $i = 1, \ldots, t-1$, $j = 0, 1, \ldots, t-1$, и вычисляет значения $y_i = P_{t-1}(x_i)$, $z_i = Q_{t-1}(x_i)$, $i = 1, \ldots, n$. Участнику A_i передается набор (x_i, y_i, z_i), $i = 1, \ldots, n$. для проверки корректности долей секрета дилер D вычисляет и помещает в общедоступном справочнике значения:

$$r_i = g^{a_i} h^{b_i} \pmod p, \quad i = 0, 1, \ldots, t-1.$$

Заметим, что (в отличие от схемы Фельдмана-Шамира) значение $r_0 = g^{S+db_0} \pmod p$ зависит от случайного числа b_0. Поэтому даже если противник и сможет вычислить значения d и $S + db_0 \pmod q$ (решив задачу дискретного логарифмирования), то это не даст ему никакой информации о секрете S.

Теперь каждый участник A_i может сделать проверку сравнения:

$$g^{y_i} h^{z_i} \equiv r_0 \, (r_1)^{x_i} \ldots (r_{t-1})^{x_i^{t-1}} \pmod p,$$

так как:

$$g^{y_i} h^{z_i} = g^{P_{t-1}(x_i)} h^{Q_{t-1}(x_i)} =$$

$$= g^S g^{a_1 x_i} \ldots g^{a_{t-1} x_i^{t-1}} h^{b_0} h^{b_1 x_i} \ldots h^{b_{t-1} x_i^{t-1}} =$$

$$= \left(g^S h^{b_0}\right) \left(g^{a_1} h^{b_1}\right)^{x_i} \ldots \left(g^{a_{t-1}} h^{b_{t-1}}\right)^{x_i^{t-1}} \equiv$$

$$\equiv r_0 \, (r_1)^{x_i} \ldots (r_{t-1})^{x_i^{t-1}} \pmod{p}.$$

Также после восстановления секрета S любой группой из t рассматриваемых участников истинность значения S можно проверить с помощью сравнения:

$$g^S h^{b_0} \equiv r_0 \pmod{p}.$$

Пусть, как и ранее, q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$, где $E_p(a,b)$ — эллиптическая кривая над полем $\mathbf{Z}_p$. Пусть некоторая точка $G \in E_p(a,b)$ имеет порядок q, $d \in \mathbb{Z}_q$ — некоторый секретный параметр, $H = [d]G$ — открытый параметр (точка эллиптической кривой).

Для разделения секрета $S \in \mathbb{Z}_q$, как и ранее, дилер D выбирает два многочлена $P_{t-1}(x)$ и $Q_{t-1}(x)$ степени $t-1$ над полем $\mathbb{Z}_q$ и вычисляет значения $y_i = P_{t-1}(x_i)$, $z_i = Q_{t-1}(x_i)$, $i = 1, \ldots, n$. Участнику A_i передается набор (x_i, y_i, z_i), $i = 1, \ldots, n$. для проверки корректности долей секрета дилер D вычисляет и помещает в общедоступном справочнике точки эллиптической кривой:

$$R_i = [a_i]G + [b_i]H, \quad i = 0, 1, \ldots, t-1.$$

Теперь каждый участник A_i может сделать проверку равенства:

$$[y_i]G + [z_i]H = R_0 + [x_i]R_1 + \ldots + [x_i^{t-1}]R_{t-1},$$

так как:

$$[y_i]G + [z_i]H = [P_{t-1}(x_i)]G + [Q_{t-1}(x_i)]H =$$

$$= [S]G + [x_i]([a_1]G) + \ldots [x_i^{t-1}]([a_{t-1}]G) +$$

$$+[b_0]H + [x_i]([b_1]H) + \ldots [x_i^{t-1}]([b_{t-1}]H) =$$

$$= ([S]G + [b_0]H) + [x_i]([a_1]G + [b_1]H) + \ldots + [x_i^{t-1}]([a_{t-1}]G + [b_{t-1}]H) =$$

$$= R_0 + [x_i]R_1 + \ldots + [x_i^{t-1}]R_{t-1}.$$

После восстановления секрета любой группой из t участников можно сделать проверку правильности восстановления секрета S :

$$R_0 = [S]G + [b_0]H.$$

13.1.4. Схема разделения секрета на основе СЛАУ

Пусть $F \ni S$ — некоторый секрет, где F — некоторое поле. Построим (n, t) пороговую схему для произвольных целых чисел n и t $(1 \leqslant t \leqslant n)$ на основе систем линейных алгебраических уравнений и теоремы Кронекера-Капелли.

В пространстве F^t выберем n векторов:

$$\overline{v}^1 = (v_1^1, v_2^1, \ldots, v_t^1),$$

$$\ldots$$

$$\overline{v}^n = (v_1^n, v_2^n, \ldots, v_t^n)$$

таким образом, чтобы любая подсистема из t векторов системы $\{\overline{v}^1, \ldots, \overline{v}^n\}$ была линейно независимой. Данное условие эквивалентно такому условию: любая матрица порядка $t \times t$, составленная из произвольных t строк матрицы:

$$\begin{pmatrix} v_1^1 & v_2^1 & \ldots & v_t^1 \\ v_1^2 & v_2^2 & \ldots & v_t^2 \\ \ldots & \ldots & \ldots & \ldots \\ v_1^n & v_2^n & \ldots & v_t^n \end{pmatrix},$$

имеет ранг t. Например, пусть $a_1, \ldots, a_t$ — попарно различные ненулевые элементы поля F. Тогда в качестве такой матрицы можно взять следующую матрицу:

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ a_1 & a_2 & \ldots & a_t \\ a_1^2 & a_2^2 & \ldots & a_t^2 \\ \ldots & \ldots & \ldots & \ldots \\ a_1^{n-1} & a_2^{n-1} & \ldots & a_t^{n-1} \end{pmatrix}.$$

Далее зафиксируем некоторым образом $t - 1$ значений:

$$x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t \in F.$$

Положим $x_i = S$ и обозначим $\overline{x} = (x_1, \ldots, x_t) \in F^t$. Вычислим следующие значения:

$$y_i = \overline{v}^i \cdot \overline{x} = v_1^i \cdot x_1 + v_2^i \cdot x_2 + \ldots + v_t^i \cdot x_t, \quad i = 1, \ldots, n.$$

Каждому участнику A_i передадим $(t+1)$-компонентный вектор $(\overline{v}^i, y_i) \in F^{t+1}$, $i = 1, \ldots, n$.

Из теоремы Кронекера-Капелли следует, что для однозначного восстановления исходного набора $\overline{x} = (x_1, \ldots, x_t)$, где $x_i = S$, достаточно t любых элементов вида $(\overline{v}^i, y_i) \in F^{t+1}$ множества $\{(\overline{v}^1, y_1), \ldots, (\overline{v}^n, y_n)\}$. При этом никакое $(t-1)$-элементное подмножество в $\{(\overline{v}^1, y_1), \ldots, (\overline{v}^n, y_n)\}$ не позволяет этого сделать.

13.1.5. Схема разделения секрета на основе равновесных двоичных кодов

Пусть имеются некоторые неотрицательные целые числа n и h, причем $0 \leqslant h \leqslant n$. Рассмотрим в пространстве:

$$V_n = \{(\varepsilon_1, \ldots, \varepsilon_n) \mid \varepsilon_1, \ldots, \varepsilon_n \in \{0, 1\}\}$$

подмножество $R(n, h)$, состоящее из всех таких элементов пространства V_n, которые имеют в своем составе ровно h единиц. Очевидно, что $|R(n, h)| = C_n^h$. Расположив все элементы множества $R(n, h)$ построчно, получим матрицу A порядка $C_n^h \times n$. Для данной матрицы верно следующее предложение.

Предложение 13.2. Пусть $1 \leqslant t \leqslant n$ и $h = n - t + 1$.

1. Выберем в матрице A произвольные t столбцов и составим из них матрицу B. Тогда каждая строка матрицы B содержит хотя бы одну единицу.

2. Если в матрице A выбрать произвольные $t - 1$ столбцов и составить из них матрицу C, то в матрице C найдется хотя бы одна строка, состоящая только из нулей.

Доказательство. 1. Каждая строка матрицы A содержит ровно $t - 1$ нулей и $n - t + 1$ единиц. Поэтому в матрице B нет нулевых строк.

2. Пусть матрица C имеет вид:

$$C = ([A]_{j_1}, [A]_{j_2}, \ldots, [A]_{j_{t-1}}),$$

где $[A]_j$ означает j-й столбец матрицы A. Из определения матрицы A следует, что в A имеется строка (с некоторым номером

i), в которой на позициях с номерами j_1, j_2,. . . ,j_{t-1} стоят нули, а все остальные значения равны единице. Тогда i-я строка матрицы C состоит только из нулей. $\square$

Пусть S — некоторый секрет. «Нарежем» его на $m = C_n^{n-t+1}$ частей k_1,. . . ,k_m. (Например, если $S \in G$, где G — некоторая аддитивная абелева группа, то можно поступить следующим образом. Сгенерируем некоторым случайным образом элементы $k_1, \ldots, k_{m-1} \in G$ и определим $k_m = S - k_1 - \ldots - k_{m-1}$. Тогда секрет S делится на m частей $k_1, \ldots, k_m$ и только наличие всех m частей позволяет восстановить секрет S.)

Далее составим матрицу A, определенную перед предложением 13.2, порядка $m \times n$. Тогда j-му участнику $(1 \leqslant j \leqslant n)$ передадим такой набор:

$$(\varepsilon_{1j} \cdot k_1, \varepsilon_{2j} \cdot k_2, \ldots, \varepsilon_{mj} \cdot k_m),$$

где $(\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{mj})^T$ — j-й столбец матрицы A.

Из предложения 13.2 видно, что любые t участников могут собрать вместе все части $k_1, \ldots, k_m$ секрета, но никакие $t-1$ участников собрать весь набор $k_1, \ldots, k_m$ не смогут.

Пример 13.1. Пусть $n = 5$, $t = 3$. Тогда:

$$h = 3, \ |R(5,3)| = C_5^3 = 10.$$

Составим матрицу A :

	1	2	3	4	5
1)	0	0	1	1	1
2)	0	1	0	1	1
3)	0	1	1	0	1
4)	0	1	1	1	0
5)	1	0	0	1	1
6)	1	0	1	0	1
7)	1	0	1	1	0
8)	1	1	0	0	1
9)	1	1	0	1	0
10)	1	1	1	0	0

Пусть $k_1, \ldots, k_{10}$ — части некоторого секрета S. Тогда каждому из 5 участников передадим такие наборы:

$$\begin{array}{ll}
1: & (0, \quad 0, \quad 0, \quad 0, \quad k_5, \ k_6, \ k_7, \ k_8, \ k_9, \ k_{10}), \\
2: & (0, \quad k_2, \ k_3, \ k_4, \ 0, \quad 0, \quad 0, \quad k_8, \ k_9, \ k_{10}), \\
3: & (k_1, \ 0, \quad k_3, \ k_4, \ 0, \quad k_6, \ k_7, \ 0, \quad 0, \quad k_{10}), \\
4: & (k_1, \ k_2, \ 0, \quad k_4, \ k_5, \ 0, \quad k_7, \ 0, \quad k_9, \ 0), \\
5: & (k_1, \ k_2, \ k_3, \ 0, \quad k_5, \ k_6, \ 0, \quad k_8, \ 0, \quad 0).
\end{array}$$

Теперь любые три участника могут собрать вместе весь набор $k_1, \ldots, k_{10}$ и тем самым восстановить секрет S, но никакие два участника этого сделать не смогут.

13.1.6. Схема разделения секрета на основе китайской теоремы об остатках

Рассмотрим (n, t) пороговую схему, где $1 < t \leqslant n$, на основе модулярной арифметики и китайской теоремы об остатках. Пусть $m_1, \ldots, m_n$ — попарно различные и попарно взаимно простые натуральные числа, которые не требуется держать в секрете. Обозначим через $\min(t)$ наименьшее произведение t различных множителей m_i :

$$\min(t) = \min_{1 \leqslant i_1 < \ldots < i_t \leqslant n} m_{i_1} \ldots m_{i_t},$$

$\max(t-1)$ — наибольшее из произведений $t-1$ различных множителей m_i :

$$\max(t-1) = \max_{1 \leqslant i_1 < \ldots < i_{t-1} \leqslant n} m_{i_1} \ldots m_{i_{t-1}}.$$

Пусть S — некоторый секрет, причем:

$$\max(t-1) < S < \min(t)$$

(любое сообщение можно преобразовать в двоичный вид, а затем нарезать на куски, каждый из которых принадлежит кольцу вычетов по модулю $\min(t) - \max(t-1) - 1$, после чего к каждому такому «куску» битов нужно добавить константу-приращение $\max(t-1)+1$, что приведет к требуемому двойному неравенству). Заметим, что $0 < S < m_1 \ldots m_n$.

Предположим, что:

$$\min(t) - \max(t-1) - 1 = H \cdot \max(t-1),$$

где H — некоторое число, причем $H \geqslant 2$. Число H будет характеризовать минимальное количество различных решений, среди которых находится секрет S, если соберутся вместе менее t участников разделения секрета.

Каждому из n участников передается своя доля секрета:

$$\alpha_i \equiv S \pmod{m_i}, \quad i = 1, \ldots, n.$$

Из следствия 1.10 видно, что для восстановления секрета S достаточно собраться вместе любым t участникам. Если же вместе собрались $t-1$ участников разделения секрета и решили систему сравнений:

$$\begin{cases} x \equiv \alpha_{i_1} \pmod{m_{i_1}}, \\ \ldots \\ x \equiv \alpha_{i_{t-1}} \pmod{m_{i_{t-1}}}, \end{cases}$$

получив при этом решение x_0, то, исходя из следствия 1.10, значение секрета S будет иметь вид

$$S = x_0 + im_{i_1} \ldots m_{i_{t-1}},$$

где $\max(t-1) < x_0 + im_{i_1} \ldots m_{i_{t-1}} < \min(t)$. Поэтому:

$$\frac{\max(t-1) - x_0}{m_{i_1} \ldots m_{i_{t-1}}} < i < \frac{\min(t) - x_0}{m_{i_1} \ldots m_{i_{t-1}}}.$$

Число же различных значений i не меньше числа:

$$\frac{\min(t) - \max(t-1) - 1}{m_{i_1} \ldots m_{i_{t-1}}} \geqslant \frac{\min(t) - \max(t-1) - 1}{\max(t-1)} \geqslant H.$$

Поэтому видно, что чем больше значение H, тем труднее перебрать все минимум H вариантов значения секрета S. Чем больше числа m_i и ближе они расположены друг к другу, тем больше значение H.

Например, если для (4,3) пороговой схемы выбрать такие модули $m_1 = 983$, $m_2 = 991$, $m_3 = 997$, $m_4 = 1000$, то в данном случае $H = 973$.

13.2. Схемы разделения секрета для произвольных структур доступа

Пусть P — конечное множество участников группы, $\widetilde{P}$ — множество, состоящее из всех возможных подмножеств множества P, R — множество, состоящее из подмножеств участников, которым разрешено восстановление секрета (правомочные коалиции), Z — множество, состоящее из подмножеств участников, которые не могут восстановить секрет (неправомочные коалиции). *Структура доступа* — разбиение $\widetilde{P} = R \cup Z$. Структура доступа обозначается как (R, Z).

Исходя из этого определения, (n, t)-пороговая схема разделения секрета — схема разделения секрета с n участниками для структуры доступа, в которой правомочными являются все коалиции, содержащие не менее t участников, а все коалиции с меньшим числом частников — неправомочны.

Структура доступа называется *монотонной*, если все надмножества правомочных коалиций также входят в R, то есть если $A \in R$, $A \subseteq B \in \widetilde{P}$, то $B \in R$.

Пусть (R, Z) — структура доступа на P. $A \in R$ называют минимальной правомочной коалицией, если $B \notin R$ всегда, когда выполнено строгое включение $B \subset A$. Множество минимальных правомочных коалиций из R обозначается как $R_{\min}$ и называется базисом R. Минимальные проавомочные коалиции однозначно задают структуру доступа. $A \in Z$ называют максимальной неправомочной коалицией, если $B \in R$ всегда, когда выполнено строгое включение $A \subset B$. Множество максимальных неправомочных коалиций из Z обозначается как $Z_{\max}$.

13.2.1. Схема Бенало-Лейхтера

Пусть задана монотонная структура доступа (R, Z) и $R_{\min}$ — множество минимальных правомочных коалиций. Пусть $R_{\min} = \{A_1, A_2, \ldots, A_m\}$. Для каждого A_i вычисляются доли секрета S для участников этого подмножества $s_{i1}, s_{i2}, \ldots, s_{i,|A_i|}$ с помо-

щью любой $(|A_i|, |A_i|)$-пороговой схемы разделения секрета. В результате каждый участник получает набор долей секрета.

Пример 13.2. Пусть S — секрет, $P = \{1, 2, 3, 4\}$, $R_{\min} = \{A_1, A_2, A_3\}$, где:

$$A_1 = \{1, 2, 3\}, \quad A_2 = \{2, 4\}, \quad A_3 = \{3, 4\}.$$

Для множества A_1 составляем $(3, 3)$-пороговую схему: по долям s_{11}, s_{12}, s_{13} однозначно восстанавливается секрет S. Аналогично, для $A_2 : s_{21}$, s_{22}; для $A_3 : s_{31}$, s_{32}. Каждый участник получает следующие доли:

$$P_1 : s_{11}, \quad P_2 : s_{12}, s_{21}, \quad P_3 : s_{13}, s_{31}, \quad P_4 : s_{22}, s_{32}.$$

Недостатком данной схемы является возрастающий объем долей секрета для каждого участника при увеличении $R_{\min}$.

13.2.2. Схема Ито-Саито-Нишизеки

Ито, Саито, Нишизеки представили так называемую технику кумулятивного массива для монотонной структуры доступа. Пусть (R, Z) — монотонная структура доступа на P, $|P| = n$, и пусть $Z_{\max} = \{B_1, B_2, \ldots, B_m\}$ — соответствующие ей максимальные неправомочные коалиции участников. Кумулятивным массивом структуры доступа (R, Z) является матрица C размера $n \times m$ над множеством $V_2 = \{0, 1\}$, где:

$$c_{ij} = \begin{cases} 0, & i \in B_j, \\ 1, & i \notin B_j, \end{cases} \qquad 1 \leqslant i \leqslant n, \ 1 \leqslant j \leqslant m.$$

В данной схеме можно использовать любую (m, m)-пороговую схему разделения секрета S и соответствующими долями секрета $s_1, s_2, \ldots, s_m$. Тогда каждый участник получает соответствующий набор:

$$P_i : c_{i1} \cdot s_1, c_{i2} \cdot s_2, \ldots, c_{im} \cdot s_m, \quad i = 1, \ldots, n.$$

Пусть $A = \{P_{i_1}, P_{i_1}, \ldots, P_{i_k}\}$ — некоторый набор участников из P. Вычеркнем в матрице C все стоки, кроме строк с номерами $i_1, i_2, \ldots, i_k$. Очевидно, что если $A \in Z$, то некоторый столбец полученной матрицы состоит из одних нулей. Если же $A \in R$, то все столбцы полученной матрицы ненулевые.

Пример 13.3. Пусть, как и в примере 13.2, S — секрет, $P = \{1, 2, 3, 4\}$, $R_{\min} = \{A_1, A_2, A_3\}$, где:

$$A_1 = \{1, 2, 3\}, \quad A_2 = \{2, 4\}, \quad A_3 = \{3, 4\}.$$

Тогда $Z_{\max} = \{B_1, B_2, B_3, B_4\}$, где:

$$B_1 = \{1, 2\}, \quad B_2 = \{1, 3\}, \quad B_3 = \{1, 4\}, \quad B_4 = \{2, 3\}.$$

Кумулятивный массив C примет такой вид:

$$C = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

Участникам из множества P передадим следующую информацию:

$$P_1 : 0, 0, 0, s_4,$$

$$P_2 : 0, s_2, s_3, 0,$$

$$P_3 : s_1, 0, s_3, 0,$$

$$P_4 : s_1, s_2, 0, s_4,$$

где s_1, s_2, s_3, s_4 — доли секрета S в $(4, 4)$-пороговой схеме.

Глава 14. Протоколы аутентификации

В традиционных (классических) криптографических системах предполагалось, что два лица, которые обмениваются секретной информацией, полностью доверяют друг другу и пытаются защитить свои сообщения от третьих лиц (перехватчиков, криптоаналитиков).

Криптография с открытым ключом значительно расширила класс задач, решаемых с помощью криптографических методов. В результате появилась потребность в интерактивных, многофазовых двусторонних обменах сообщениями между участниками, которые не всегда доверяют друг другу, в передаче информации между несколькими участниками. Криптографический протокол — протокол, предназначенный для выполнения функций криптографической системы, в процессе выполнения которого участники используют криптографические алгоритмы.

Протокол аутентификации — протокол установления подлинности сторон, участвующих во взаимодействии, но не доверяющих друг другу. Одной из основных целей протоколов аутентификации является обеспечение контроля доступа к определенным ресурсам, таким как банковские счета, базы данных, здания, сооружения и т.д. Протокол аутентификации включает двух участников: доказывающего участника A, проходящего аутентификацию, и проверяющего участника B, проверяющего аутентичность доказывающего. Целью протокола является установление того, что проверяемый действительно является участник A. Различают протоколы аутентификации с односторонней и взаимной аутентификацией.

Идентификация — процедура установления присвоенного данной стороне уникального системного имени — идентифика-

тора, которое позволяет отличать ее от других сторон. Обычно процедура идентификации заключается в предъявлении этого имени и предшествует процедуре аутентификации, то есть подтверждению правильности идентификации. Термин идентификация часто для краткости используют для обозначения общей процедуры идентификации/аутентификации сторон.

Протоколы аутентификации разделяют на следующие классы:

- протоколы, основанные на паролях (слабая аутентификация);
- протоколы, использующие технику «запрос–ответ» (сильная аутентификация);
- протоколы, основанные на технике доказательства знания;
- протоколы доказательства знания с нулевым разглашением.

14.1. Протоколы аутентификации, использующие пароли (слабая аутентификация)

Частным вариантом использования процедуры простой аутентификации является парольная защита входа в компьютерную систему [27]. Например, пользователь формирует некоторую случайную информацию и, сохраняя ее в секрете, использует как пароль. Пароль в явном виде не хранится в памяти ЭВМ (или другого устройства, применяемого для выполнения аутентификации). Это требование направлено на то, чтобы потенциальный внутренний нарушитель не имел возможности извлечь из машинной памяти чужой пароль и присвоить себе полномочия другого пользователя. Для того, чтобы система защиты могла идентифицировать легальных (санкционированных) пользователей, в памяти ЭВМ хранятся образы их паролей, вычисленные по специальному криптографическому алгоритму, реализующему, так называемую, одностороннюю функцию $y = f(x)$, где f криптографический алгоритм — односторонняя функция. Основное требование к односторонней

функции состоит в том, чтобы сложность вычисления значения функции по аргументу (по входу) была низкой, а сложность определения значения аргумента, для которого значение функции (значение выхода) было бы равно случайно выбранному значению, была высокой (неосуществимой за обозримое время при использовании всех вычислительных ресурсов доступных предполагаемому нарушителю).

Аутентификация пользователя на рабочей станции может быть выполнена следующим образом:

1. Запрос на ввод идентификатора со стороны системы защиты.
2. Ввод пользователем своего идентификатора (имени) name.
3. Запрос на ввод пароля со стороны системы защиты.
4. Ввод пользователем пароля password.
5. Вычисление системой защиты значения односторонней функции y, соответствующей значению аргумента $x = password$.
6. Сравнение системой защиты значения $f(password)$ со значением образа (s) пароля, соответствующего пользователю с идентификатором name.

Если $f(password) = s$, то система защиты предоставляет пользователю права доступа (полномочия), соответствующие идентификатору name. В противном случае в журнале учета работы пользователей регистрируется событие попытки несанкционированного доступа. Для того, чтобы выдать себя за санкционированного пользователя нарушитель должен ввести правильный пароль. Зная образ s, вычислительно невозможно определить пароль password. Если в системе защиты предусмотрены механизмы противодействия перехвату пароля с помощью программных или аппаратных закладок, а также через побочные электромагнитные излучения и наводки (ПЭМИН), акустический и оптический канал, то данный способ аутентификации

пользователей обеспечивает высокую надежность защиты от узурпирования чужих полномочий. Рассмотренный пример относится к аутентификации пользователей на рабочих станциях, т.е. к задаче защиты входа в ЭВМ.

14.2. Протоколы аутентификации, использующие технику «запрос–ответ» (сильная аутентификация)

Идея построения криптографических протоколов аутентификации типа «запрос–ответ» состоит в том, что доказывающий убеждает проверяющего в своей аутентичности путем демонстрации знания некоторого секрета без предъявления самого секрета. В таких протоколах используются случайные числа или метки времени. Метки времени используют для обеспечения гарантий своевременности и единственности сообщений, а также для обнаружения попыток навязывания ранее переданной информации. Сообщение принимается, если его временна́я метка находится в пределах так называемого «временно́го окна».

14.2.1. «Запрос–ответ» с использованием симметричных алгоритмов шифрования

Для реализации данных протоколов необходимо, чтобы доказывающий и проверяющий имели общий секретный ключ. Введем следующие обозначения: r_A — случайное число, вырабатываемое абонентом A, t_A — временна́я метка A, k_{AB} — общий секретный ключ A и B, E_k — алгоритм шифрования на ключе k, A — идентификатор участника A (для краткости будем использовать тот же символ).

Рассмотрим некоторые примеры протоколов.

1. *Односторонняя аутентификация с использованием временно́й метки.* Доказывающий A передает проверяющему B свою временну́ю метку и идентификатор, зашифрованный на общем ключе:

$$A \to B: \ E_{k_{AB}}(t_A, B).$$

Проверяющий B расшифровывает данное сообщение, проверяет соответствие допустимому интервалу временно́й метки и совпадение полученного и собственного идентификаторов. Это необходимо для того, чтобы не позволить противнику немедленно переадресовать сообщение абоненту A.

2. *Односторонняя аутентификация с использованием случайных чисел.* В данном случае проверяющий B, расшифровав сообщение, проверяет, соответствует ли полученное число случайному числу, отправленное им ранее абоненту A. Также B проверяет правильность полученного идентификатора:

$$\begin{aligned} &1)\ \ A \leftarrow B:\ r_B; \\ &2)\ \ A \rightarrow B:\ E_{k_{AB}}(r_B, B). \end{aligned}$$

Для предотвращения возможности криптоанализа алгоритма E_k абонент A может во второе сообщение ввести свое случайное число.

3. *Взаимная аутентификация с использованием случайных чисел.* Данный протокол имеет вид:

$$\begin{aligned} &1)\ \ A \leftarrow B:\ r_B; \\ &2)\ \ A \rightarrow B:\ E_{k_{AB}}(r_A, r_B, B); \\ &3)\ \ A \leftarrow B:\ E_{k_{AB}}(r_A, r_B). \end{aligned}$$

Отличие этого протокола от предыдущего в том, что здесь каждый из участников поочередно выполняет роли проверяющего и претендента, т.е. они проверяют аутентичность друг друга. Подобного рода протоколы с силу их симметричности называют *протоколами рукопожатия.*

Эти три конструкции реализованы в рассматриваемых далее протоколах, предложенных в качестве стандартов Международной организацией по стандартизации (International Organization for Standardization — ISO).

Протокол односторонней аутентификации:

$$\begin{aligned} &1)\ \ A \leftarrow B:\ r_B,\ M_1; \\ &2)\ \ A \rightarrow B:\ M_3,\ E_{k_{AB}}(r_B, B, M_2). \end{aligned}$$

Поля $M_1 - M_2$ могут содержать произвольные текстовые сообщения.

Протокол взаимной аутентификации двухпроходный:

$$1)\ \ A \to B:\ \ M_2,\ \ E_{k_{AB}}([t_A|r_A], B, M_1);$$
$$2)\ \ A \leftarrow B:\ \ M_4,\ \ E_{k_{AB}}([t_B|r_B], A, M_3).$$

В данном протоколе выбор случайного числа r_A или временно́й метки t_A зависит от конкретной реализации.

Протокол взаимной аутентификации трехпроходный:

$$1)\ \ A \leftarrow B:\ \ r_B,\ \ M_1;$$
$$2)\ \ A \to B:\ \ M_3,\ \ E_{k_{AB}}(r_A, r_B, B, M_2);$$
$$3)\ \ A \leftarrow B:\ \ M_5,\ \ E_{k_{AB}}(r_B, r_A, M_4).$$

14.2.2. «Запрос–ответ» с использованием асимметричных алгоритмов шифрования

В протоколе аутентификации, построенном на основе шифрования с открытым ключом, доказывающий может продемонстрировать владение секретным ключом одним из двух способов: путем расшифрования запроса, зашифрованного на его открытом ключе; путем добавления к запросу своей электронной подписи. Рассмотрим эти способы более подробно.

Протоколы аутентификации, не использующие электронную подпись

Протокол односторонней аутентификации. Пусть h — некоторая однонаправленная функция; E_A, D_A — алгоритмы соответственно зашифрования и расшифрования абонента A. Первый способ основан на следующем протоколе:

$$1)\ \ A \leftarrow B:\ \ h(r),\ \ B,\ \ E_A(r, B);$$
$$2)\ \ A \to B:\ \ r.$$

Проверяющий B выбирает случайное число r, вычисляет значение $h(r)$ и запрос $E_A(r, B)$. Доказывающий расшифровывает запрос $E_A(r, B)$ и проверяет совпадение значений хеш-функции и идентификаторов. Проверяющий B идентифицирует A, если полученное число r совпадет с имеющимся у него числом.

Протокол NSPK (Needham-Schroeder-Public Key Protocol). Данный протокол является протоколом взаимной аутентификации, предложенный Р.М. Нидхэмом и М.Д. Шредером:

$$\begin{array}{ll} 1) & A \to B : E_B(r_A, A); \\ 2) & A \leftarrow B : E_A(r_A, r_B); \\ 3) & A \to B : E_B(r_B). \end{array}$$

Заметим, что данный протокол имеет уязвимости.

Протоколы аутентификации, использующие электронную подпись

Пусть r_A — случайное число доказывающего A, t_A — его метка времени, S_A — алгоритм электронной подписи A. Предполагается, что алгоритм проверки электронной подписи доказывающего известен проверяющему. Пусть $cert_A$ — сертификат открытого ключа для алгоритма проверки электронной подписи A, pk_A — открытый ключ абонента A. В большинстве случаев сертификат имеет вид:

$$cert_A = (pk_A, A, S_T(pk_A, A)),$$

где S_T — электронная подпись центра сертификации T.

Для аутентификации могут быть использованы три протокола.

1. Односторонняя аутентификация с использованием временных меток:

$$A \to B : \ cert_A, \ t_A, \ B, \ S_A(t_A, B).$$

Получив сообщение, B проверяет, что временная метка находится в допустимом интервале, идентификатор B совпадает с его собственным и то, что электронная подпись под этими двумя полями верна.

2. Односторонняя аутентификация с использованием случайных чисел:

$$\begin{array}{ll} 1) & A \leftarrow B : \ r_B; \\ 2) & A \to B : \ cert_A, \ r_A, \ B, \ S_A(r_A, r_B, B). \end{array}$$

Получив сообщение, B проверяет, что идентификатор B соответствует его идентификатору и что электронная подпись под строкой (r_A, r_B, B) верна.

3. Взаимная аутентификация с использованием случайных чисел:

$$1)\ A \leftarrow B:\ r_B;$$
$$2)\ A \rightarrow B:\ cert_A,\ r_A,\ B,\ S_A(r_A, r_B, B);$$
$$3)\ A \leftarrow B:\ cert_B,\ A,\ S_B(r_B, r_A, A).$$

Рассмотрим примеры таких протоколов, предложенных Международной организацией стандартов ISO.

Протокол односторонней аутентификации ISO1. Пусть участник A имеет открытый pk_A и секретный sk_A ключи для зашифрования E_A и расшифрования D_A соответственно. Аутентификация проводится с использованием случайного числа r_A :

$$A \rightarrow B:\ cert_A,\ r_A,\ B,\ M,\ D_A(r_A, B, M),$$

где M — произвольное сообщение, $cert_A = (pk_A, A, D_T(pk_A, A))$. Здесь в качестве электронной подписи выступает подпись на основе асимметричного шифрования.

Протокол взаимной аутентификации двухпроходный ISO2. Он является двукратным повторением протокола ISO1:

$$1)\ A \rightarrow B:\ cert_A,\ r_A,\ B,\ M_2,\ D_A(r_A, B, M_1);$$
$$2)\ A \leftarrow B:\ cert_B,\ r_B,\ A,\ M_4,\ D_B(r_B, A, M_3),$$

где M_1–M_4 — произвольные текстовые сообщения.

Протокол взаимной аутентификации трехпроходный ISO3:

$$1)\ A \leftarrow B:\ r_B,\ M_1;$$
$$2)\ A \rightarrow B:\ cert_A,\ r_A,\ r_B,\ B,\ M_3,\ D_A(r_A, r_B, B, M_2);$$
$$3)\ A \leftarrow B:\ cert_B,\ r_B,\ r_A,\ A,\ M_5,\ D_B(r_B, r_A, A, M_4).$$

14.3. Протоколы аутентификации с нулевым разглашением знания

В парольных схемах противник может запомнить передаваемые сообщения и в следующий раз использовать эту информа-

цию. В протоколах типа «запрос–ответ» противник, контролируя канал связи, может навязывать специально подобранные запросы и, анализируя ответы, получить информацию о секрете. Чтобы избежать этого, применяют протоколы *доказательства знания* (некоторой секретной информации), которые обладают дополнительным свойством нулевого разглашения секрета.

Доказательство знания — интерактивное доказательство, при котором доказывающий убеждает проверяющего в том, что он владеет секретной информацией, не раскрывая ее.

Интерактивное доказательство — понятие теории сложности вычислений, составляющее основу понятия доказательства с нулевым разглашением. Интерактивное доказательство — доказательство путем выполнения протокола с двумя участниками, доказывающим и проверяющим, в процессе работы которых участники обмениваются сообщениями (запросы и ответы), обычно зависящими от случайных чисел, которые могут содержаться в секрете. Цель доказывающего — убедить проверяющего в истинности некоторого утверждения. Проверяющий либо принимает, либо отвергает доказательство. В отличие от обычного математического понятия доказательства в данном случае доказательство носит не абсолютный, а вероятностный характер и характеризуется двумя вероятностями. Если доказываемое утверждение верно, то доказательство должно быть верным с вероятностью, стремящейся к единице при увеличении числа циклов протокола. Если же доказываемое утверждение ложно, то при увеличении числа циклов протокола вероятность правильности доказательства должна стремиться к нулю. Криптографическое качество протокола интерактивного доказательства характеризуется свойствами *полноты*, *корректности* и *нулевого разглашения*.

Полнота — свойство криптографического протокола, означающее, что при выполнении честными участниками протокол решает ту задачу, для которой он создан.

Корректность — способность криптографического протокола противостоять угрозам со стороны противника и/или нару-

шителя, не располагающего необходимой секретной информацией, но пытающегося выполнить протокол за участника, который по определению должен такой информацией владеть.

Нулевое разглашение — свойство протокола доказательства знания, обеспечивающее такое его выполнение, что никакая информация о доказываемом утверждении, кроме факта его истинности, не может быть получена нечестным проверяющим из переданных сообщений за время полиномиально зависящее от суммарной длины этих сообщений.

Протокол с нулевым разглашением — синоним понятия «доказательство знания с нулевым разглашением».

Протоколы аутентификации, построенные на основе протоколов с нулевым разглашением знания, используют два ключа (секретный и открытый), и применяются для аутентификации удаленных абонентов. На самом деле в некоторых таких протоколах аутентификации присутствует изначальная утечка информации о секретном ключе по значению открытого ключа. Данная утечка является допустимой, так как вычисление секретного ключа по открытому ключу является вычислительно трудной задачей. Эта вычислительная трудность задачи принимается за стойкость протокола аутентификации. Например, протокол аутентификации Шнорра основан на трудной задаче дискретного логарифмирования. Когда говорится о том, что протокол аутентификации построен на основе протокола с нулевым разглашением секрета, то имеется ввиду, что в ходе протокола не происходит никакой дополнительной утечки информации о секрете. Существуют также протоколы аутентификации на основе техники доказательстве знания, построенные на основе NP-полных задач. Такими протоколами, в частности, являются протокол аутентификации на основе доказательства изоморфизма графов, протокол аутентификации на основе задачи о нахождении гамильтонова цикла в графе и т.д. Эти протоколы основаны на NP-полных задачах, поэтому являются независимыми от квантовых вычислений.

14.3.1. Протокол аутентификации Фиата-Шамира

На основе электронной подписи Фиата-Шамира, основанной на сложности задачи извлечения квадратного корня по модулю большого составного числа n с неизвестным разложением на множители, построим протокол аутентификации.

Доверенный центр T фиксирует два различных простых числа p и q, вычисляет $n = pq$. Число n доверенный центр сообщает всем доказывающим. Числа p и q остаются в секрете или уничтожаются. Доверенный центр необходим для того, чтобы задача извлечения корня была сложной.

В качестве секретного ключа доказывающий участник A выбирает случайное число $x \in \mathbb{Z}_n \setminus \{0\}$, $(x, n) = 1$, открытым ключом объявляет значение $y = (x^{-1})^2 \pmod{n}$. Протокол имеет следующий вид:

1. Доказывающий A выбирает случайное число k, для которого выполнено неравенство $1 \leqslant k \leqslant n - 1$, и отправляет участнику B число $r = k^2 \pmod{n}$.
2. Проверяющий B отвечает случайным запросом a, $a \in \{0, 1\}$.
3. В ответ участник A посылает проверяющему B значение $s = kx^a \pmod{n}$.
4. Теперь участник B проверяет сравнение $s^2 y^a \equiv r \pmod{n}$.

Эти четыре шага повторяются независимо t раз, причем проверяющий B принимает доказательство владения участником A секрета x, если все эти итерации приводят к положительному ответу.

Полнота. Доказывающий A знает значение секрета x, поэтому он в состоянии ответить на оба вопроса. Проверяющий B убеждается в справедливости соотношений:

$$s^2 y^a = k^2 x^{2a} (x^{-1})^{2a} \equiv k^2 \equiv r \pmod{n}.$$

Корректность. Наличие запроса a необходимо, чтобы доказывающий A был в состоянии ответить на любой из двух вопросов, ответ на один из которых требует знание секрета x, а ответ

на другой предотвращает попытку обмана. Противник не знает секрет x, поэтому он может подготовиться к ответу только на один из вопросов. Если он подготовится к первому вопросу, выбрав произвольное число z и передав проверяющему B число $z^2 \pmod n$, то на запрос $a = 0$, направит правильный ответ $s = z$. Но при этом он не сможет ответить на второй вопрос. Также противник может ответить на второй вопрос следующим образом. Противник может зафиксировать произвольное число z и отправить первым сообщением значение $r = z^2 y \pmod n$, а вторым сообщением на запрос $a = 1$ значение $s = z \pmod n$. Тогда $s^2 y^a \equiv r \pmod n$, но противник не сможет при этом ответить на первый вопрос. Таким образом, вероятность обмана при t-кратной итерации не превосходит значения 2^{-t}. Если $40 \leqslant t \leqslant 160$, то вероятность обмана можно снизить до приемлемо низкой величины.

Нулевое разглашение. На запрос проверяющего B $a = 0$ ответ $s = r$ не несет никакой информации о секретном ключе x. Ответ $s = kx$ на запрос $a = 1$ также не несет информации о секрете x, так как зависит от случайного числа k, не известное проверяющему B.

Опишем следующий формальный прием доказательства свойства нулевого разглашения. Покажем, что проверяющий B мог самостоятельно получить правильную тройку сообщений протокола (r, a, s). Проверяющий B мог сгенерировать случайные числа a, s, $1 \leqslant s \leqslant n - 1$, и вычислить $r = s^2 y^a \pmod n$. В этом случае тройка (r, a, s) является корректной тройкой сообщений, которая может появиться при реализации протокола. При этом эта тройка получена без участия абонента A. Так как таким способом может быть получена любая тройка сообщений протокола, то множество допустимых троек протокола совпадает с множеством троек, которые проверяющий B может получить и сам. Поэтому при наличии доказательства знания секрета x от абонента A в виде таких троек, которые проверяющий B может получить и сам, он не может получить из них дополнительную информацию о секрете x участника A.

14.3.2. Протокол Фейга-Фиата-Шамира

Достоинством многораундового протокола Фиата-Шамира является его сравнительно низкая вычислительная сложность — каждая из сторон участвующих в протоколе выполняет не более $2t$ модульных умножений, где t — заданное число раундов (итераций). Однако, существенным недостатком всех многораундовых протоколов является необходимость выполнения очень большого числа чередующихся пересылок сообщений от доказывающего к проверяющему и обратно. Этот недостаток можно устранить, используя механизм объединения всех случайных однобитовых запросов проверяющего в единую случайную битовую строку, которая направляется доказывающему целиком, и свертки всех ответов в единое значение, которое направляется от доказывающего к проверяющему.

Данный протокол является модификацией протокола аутентификации Фиата-Шамира. Здесь используется не один, а целый набор секретных ключей.

Доверенный центр T фиксирует два различных простых числа p и q, вычисляет $n = pq$. Число n доверенный центр сообщает всем доказывающим. Числа p и q остаются в секрете.

В качестве секретного ключа доказывающий участник A выбирает случайные числа $x_i \in \mathbb{Z}_n \setminus \{0\}$, $(x_i, n) = 1$, $i = 1, \ldots, m$, открытыми ключами объявляет значения $y_i = (x_i^{-1})^2 \pmod n$, $i = 1, \ldots, m$. Протокол имеет следующий вид:

1. Доказывающий A выбирает случайное число k, для которого выполнено $1 \leqslant k \leqslant n-1$, и отправляет участнику B число $r = k^2 \pmod n$.

2. Проверяющий B отвечает случайным запросом:
$$(a_1, \ldots, a_m) \in \{0,1\}^m.$$

3. В ответ участник A посылает проверяющему B значение $s = kx_1^{a_1} \ldots x_m^{a_m} \pmod n$.

4. Теперь участник B проверяет сравнение:
$$s^2 y_1^{a_1} \ldots y_m^{a_m} \equiv r \pmod n.$$

Эти четыре шага повторяются независимо t раз.

В данном протоколе вероятность ошибки проверяющего в t проходах цикла не превосходит значения 2^{-mt}. Авторы протокола рекомендовали выбирать $t = 4$, $m = 5$.

Доказательство полноты, корректности и нулевого разглашения аналогично, как и в предыдущем случае.

14.3.3. Протокол аутентификации с нулевым разглашением без доверенного центра

Рассмотрим итеративный и трехпроходный протоколы аутентификации, предложенные в работе [19].

Итеративный протокол. Для устранения необходимости наличия в протоколе доверенного центра можно предложить использовать трудность задачи извлечения корней большой простой степени по простому модулю со специальной структурой, а именно простое число p, имеющее структуру $p = Nq^2 + 1$, где разрядность числа q 160 бит и разрядность числа N не менее 864 бит. В протоколе, основанном на трудности извлечения корней q-й степени по модулю p, абонент A выбирает случайное число x, $1 \leqslant x \leqslant p-1$, и вычисляет значение своего открытого ключа $y = x^{-q} \pmod p$.

Протокол состоит из t-кратного повторения следующих шагов:

1. Доказывающий A выбирает случайное число k, для которого выполнено неравенство $1 \leqslant k \leqslant p-1$, и отправляет участнику B число $r = k^q \pmod p$.
2. Проверяющий B отвечает случайным запросом a, $a \in \{0, 1\}$.
3. В ответ участник A посылает проверяющему B значение $s = kx^a \pmod p$.
4. Теперь участник B проверяет сравнение $s^q y^a \equiv r \pmod p$.

Полнота. Доказывающий A знает значение секрета x, поэтому он в состоянии ответить на оба вопроса. Проверяющий B

убеждается в справедливости соотношений:

$$s^q y^a = k^q x^{aq} (x^{-q})^a \equiv k^q \equiv r \pmod{p}.$$

Корректность. Противник может подготовиться к запросу $a = 0$. В этом случае он сгенерирует случайное целое число k, $1 \leqslant k \leqslant p-1$, вычислит $r = k^q \pmod{p}$ и отправит проверяющему B значение r. Тогда на запрос $a = 0$ противник передаст проверяющему B значение $s = k \pmod{p}$. В этом случае он не сможет ответить правильно на запрос $a = 1$. Так же противник может подготовиться к запросу $a = 1$. В этом случае он сгенерирует случайное число s, $1 \leqslant s \leqslant p-1$, вычислит $r = s^q y \pmod{p}$. На первом шаге протокола противник передаст значение r, а на запрос $a = 1$ передаст s. Поэтому он верно ответит на запрос $a = 1$, но не сможет ответить на запрос $a = 0$, так как придется решить трудную задачу извлечения корней большой простой степени по простому модулю. Таким образом, вероятность обмана при t-кратной итерации не превосходит значения 2^{-t}.

Нулевое разглашение. Проверяющий B может самостоятельно получить любую тройку сообщений протокола (r, a, s). Для этого он может генерировать значения s, a, $1 \leqslant s \leqslant p-1$, $a \in \{0, 1\}$, и вычислять $r = s^q y^a \pmod{p}$. Такие тройки являются корректными тройками протокола и получены без участия абонента A. При этом множество всех корректных троек протокола, получаемых таким образом проверяющим B, совпадают с множеством всех допустимых троек протокола. Поэтому при наличии доказательства знания секрета x от абонента A в виде таких троек, которые проверяющий B может получить и сам, он не может получить из них дополнительную информацию о секрете x участника A.

Трехпроходный протокол. Рассмотрим реализацию трехпроходного протокола на основе предыдущего итеративного протокола, использующего простые числа вида $p = Nq^2 + 1$. Абонент A выбирает случайное секретное число x_i, $1 \leqslant x_i \leqslant p-1$, и вычисляет значение открытого ключа $y_i = x_i^{-q} \pmod{p}$, $i = 1, \ldots, m$.

Трехпроходный протокол аутентификации Фиата-Шамира имеет следующий вид:

1. Доказывающий A генерирует случайное целое число k, где $1 \leqslant k \leqslant q-2$, вычисляет $r = k^q \pmod p$ и отправляет проверяющему B значение r.

2. Проверяющий B генерирует случайную битовую строку $(a_1, \ldots, a_m) \in \{0,1\}^m$.

3. Абонент A вычисляет и передает проверяющему B значение:
$$s = k \cdot \prod_{i=1}^{m} x_i^{a_i} \pmod p.$$

4. Теперь если выполнено равенство:
$$r = s^q \prod_{i=1}^{m} y_i^{a_i} \pmod p,$$
то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Доказательство полноты, корректности и нулевого разглашения аналогично, как и в предыдущем случае.

Вероятность обмана проверяющего со стороны нарушителя, пытающегося выдать себя за подлинного владельца открытого ключа, равна 2^{-m}.

14.3.4. Протокол аутентификации Шнорра

Данный протокол получается простой модификацией алгоритма формирования электронной подписи Шнорра, в котором вместо подписываемого сообщения выступает запрос a. Он основан на трудной задаче дискретного логарифмирования.

Пусть p — простое число, q — простой делитель числа $p-1$, $g \in \mathbb{Z}_p$, имеющий порядок q. Абонент A выбирает случайное число x, для которого выполнено $1 \leqslant x \leqslant q-2$, и вычисляет значение $y = g^{-x} \pmod p$.

Число x является секретным ключом, набор (p, q, g, y) — открытым ключом. Пусть $cert_A$ — сертификат открытого ключа абонента A.

Протокол аутентификации Шнорра имеет следующий вид:

1. Доказывающий A генерирует случайное целое число k, где $1 \leqslant k \leqslant q-2$, вычисляет $r = g^k \pmod p$ и отправляет проверяющему B $cert_A$ и значение r.

2. Проверяющий B генерирует случайное число a (с условием $0 \leqslant a \leqslant 2^t - 1$, где t — некоторый параметр), которое передает абоненту A.

3. Абонент A вычисляет и передает проверяющему B значение $s = k + ax \pmod q$.

4. Теперь если выполнено равенство $r = g^s y^a \pmod p$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Заметим, что в первоначальном варианте число a от проверяющего B должно состоять из некоторого числа бит (например, 100 бит). Однако лучше использовать битовые запросы $a \in \{0, 1\}$ и повторять весь трехшаговый цикл некоторое число t раз.

Предложение 14.1. Если противник перехватит тройки протокола вида (r, a_1, s_1), (r, a_2, s_2), $a_1 \neq a_2$, то он может (с помощью обобщенного алгоритма Евклида) найти значение секретного ключа x абонента A.

Доказательство. Исходя из равенств:

$$r = g^{s_1} y^{a_1} \pmod p, \quad r = g^{s_2} y^{a_2} \pmod p,$$

получаем:

$$g^{s_1 - s_2} \equiv y^{a_2 - a_1} \equiv g^{x(a_1 - a_2)} \pmod p.$$

Так как $g^q \equiv 1 \pmod p$, то:

$$x = -\log_g y = (s_1 - s_2)(a_1 - a_2)^{-1} \pmod q. \qquad \square$$

Из данного предложения следует, что генерация случайных чисел k в протоколе должна выполняться очень качественным генератором.

Предложение 14.2. Если для некоторых r противник может по значениям a вычислять значения s, фигурируемых в протоколе Шнорра, то он может использовать данный алгоритм для вычисления дискретных логарифмов.

Доказательство следует из предложения 14.1. □

Полнота. Доказывающий A знает значение секрета x, поэтому он в состоянии ответить на оба вопроса. Проверяющий B убеждается в справедливости соотношений:

$$g^s y^a = g^k g^{ax} (g^{-x})^a \equiv g^k \equiv r \pmod{p}.$$

Корректность. Противник не знает секретного ключа x, поэтому для случая больших чисел a от B вероятность обмануть проверяющего B не превосходит 2^{-t}. Рассмотрим случай, когда $a \in \{0, 1\}$. Противник может подготовиться к запросу $a = 0$. В этом случае он сгенерирует случайное целое число k, $1 \leqslant k \leqslant q - 2$, вычислит $r = g^k \pmod{p}$ и отправит проверяющему B значение r. Тогда на запрос $a = 0$ противник передаст проверяющему B значение $s = k \pmod{q}$. В этом случае он не сможет ответить правильно на запрос $a = 1$. Так же противник может подготовиться к запросу $a = 1$. В этом случае он сгенерирует случайное число s, $1 \leqslant s \leqslant q - 1$, вычислит $r = g^s y \pmod{p}$. На первом шаге противник передаст значение r, а на запрос $a = 1$ передаст s. Поэтому он верно ответит на запрос $a = 1$, но не сможет ответить на запрос $a = 0$, так как придется решить трудную задачу дискретного логарифмирования и вычислить $\log_g r$. С вероятностью $1/q$ противник может только попытаться угадать значение $\log_g r$. Поэтому по формуле полной вероятности вероятность успешной аутентификации на каждом из t шагов не превышает значения $\frac{1}{2}(1 + \frac{1}{q})$.

Нулевое разглашение. Проверяющий B может самостоятельно получить любую тройку сообщений протокола (r, a, s). Для

этого он может генерировать значения s, a, $1 \leqslant s \leqslant q-1$, $1 \leqslant a \leqslant q-1$, и вычислять $r = g^s y^a \pmod p$. Такие тройки являются корректными тройками протокола и получены без участия абонента A. При этом множество всех корректных троек протокола, получаемых таким образом проверяющим B, совпадают с множеством всех допустимых троек протокола. Поэтому при наличии доказательства знания секрета x от абонента A в виде таких троек, которые проверяющий B может получить и сам, он не может получить из них дополнительную информацию о секрете x участника A.

14.3.5. Трехпроходный протокол аутентификации Шнорра

В данном случае используется не t итераций протокола Шнорра, а только один раунд с использованием битовой строки запроса.

Пусть p — простое число, q — простой делитель числа $p-1$, $g \in \mathbb{Z}_p$, имеющий порядок q. Абонент A выбирает случайные числа x_i, $1 \leqslant x_i \leqslant q-2$, и вычисляет значения $y_i = g^{-x_i} \pmod p$, $i = 1, \ldots, m$.

Числа $x_1, \ldots, x_m$ являются секретными ключами, набор p, q, g, $y_1, \ldots, y_m$ — открытым ключом. Пусть $cert_A$ — сертификат открытого ключа абонента A.

Трехпроходный протокол аутентификации Шнорра имеет следующий вид:

1. Доказывающий A генерирует случайное целое число k, где $1 \leqslant k \leqslant q-2$, вычисляет $r = g^k \pmod p$ и отправляет проверяющему B значение r.

2. Проверяющий B генерирует случайную битовую строку $(a_1, \ldots, a_m) \in \{0,1\}^m$ и передает ее абоненту A.

3. Абонент A вычисляет и передает проверяющему B значение:
$$s = k + \sum_{i=1}^{m} a_i x_i \pmod q.$$

4. Теперь если выполнено равенство:

$$r = g^s \prod_{i=1}^{m} y_i^{a_i} \pmod{p},$$

то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Вероятность обмана проверяющего со стороны нарушителя, пытающегося выдать себя за подлинного владельца открытого ключа, равна 2^{-m}.

14.3.6. Протокол аутентификации Окамото

Данный протокол является модификацией протокола Шнорра.

Пусть, как и в предыдущем случае, p — простое число, q — простой делитель числа $p-1$, $g_1, g_2 \in \mathbb{Z}_p$, имеющие порядок q. Абонент A выбирает пару случайных чисел x_1, x_2, для которых выполнены неравенства $1 \leqslant x_1 \leqslant q-2$, $1 \leqslant x_2 \leqslant q-2$, и вычисляет значение $y = g_1^{-x_1} g_2^{-x_2} \pmod{p}$.

Пара чисел x_1, x_2 является секретным ключом, набор p, q, g_1, g_2, y — открытым ключом. Пусть $cert_A$ — сертификат открытого ключа абонента A.

Протокол аутентификации Окамото имеет следующий вид:

1. Доказывающий A генерирует случайные целые числа k_1, k_2, $1 \leqslant k_1 \leqslant q-2$, $1 \leqslant k_2 \leqslant q-2$, вычисляет $r = g_1^{k_1} g_2^{k_2} \pmod{p}$ и отправляет проверяющему B $cert_A$ и значение r.

2. Проверяющий B генерирует случайное число a (с условием $0 \leqslant a \leqslant 2^t < q-1$), которое передает абоненту A.

3. Доказывающий A вычисляет и передает проверяющему B значения:

$$\begin{cases} s_1 = k_1 + ax_1 \pmod{q}, \\ s_2 = k_2 + ax_2 \pmod{q}. \end{cases}$$

4. Если выполнено равенство $r = g_1^{s_1} g_2^{s_2} y^a \pmod{p}$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Полнота. Доказывающий A знает значение секрета x_1, x_2, поэтому он в состоянии ответить на оба вопроса. Проверяющий B убеждается в справедливости соотношений:

$$g_1^{s_1} g_2^{s_2} y^a = g_1^{k_1} g_1^{ax_1} g_2^{k_2} g_1^{ax_1} (g_1^{-x_1})^a (g_2^{-x_2})^a \equiv g_1^{k_1} g_2^{k_2} \equiv r \pmod{p}.$$

Доказательство *корректности* и *нулевого разглашения* для случая $a \in \{0, 1\}$ провести самостоятельно.

14.3.7. Модификация (усиление) протокола Шнорра

В 1990 году Жарк Жиральт на конференции EUROCRYPT [58] предложил в схеме аутентификации Шнорра использовать составной модуль вида $n = pq$, где числа p и q имеют вид $p = 2d\widetilde{p} + 1$, $q = 2d\widetilde{q} + 1$, где d, $\widetilde{p}$, $\widetilde{q}$ — простые числа. Пусть g — такое целое число, которое имеет порядок d по модулю p и по модулю q, x — секретный ключ, $0 < x < d$, $y = g^{-x} \pmod{n}$ — открытый ключ.

Модифицированный протокол Шнорра имеет следующий вид:

1. Доказывающий A генерирует случайное целое число k, где $1 \leqslant k \leqslant d - 2$, вычисляет $r = g^k \pmod{n}$ и отправляет проверяющему B $cert_A$ и значение r.

2. Проверяющий B генерирует случайное число a (с условием $0 \leqslant a \leqslant 2^t - 1$, где t — некоторый параметр), которое передает абоненту A.

3. Абонент A вычисляет и передает проверяющему B значение $s = k + ax \pmod{d}$.

4. Теперь если выполнено равенство $r = g^s y^a \pmod{n}$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

В рассматриваемой работе предлагается использовать параметры следующей длины: $|d|$=200 бит, $|\widetilde{p}|$=$|\widetilde{q}|$=300 бит, $|n|$ = 1000 бит.

Для взлома полученной системы аутентификации потребуется решить задачу факторизации модулю n и задачу дискретного логарифмирования по модулям p и q.

14.3.8. Модификация протоколов Шнорра и Окамото на эллиптических кривых

Рассмотрим модификации протоколов Шнорра и Окамото на эллиптических кривых, предложенные в работе [30].

Модификация протокола Шнорра на эллиптических кривых.

Пусть — эллиптическая кривая, известная участникам информационного процесса, G — предварительно согласованная и опубликованная точка порядка n этой кривой. Абонент выбирает секретное число x, $1 < x < n$, и вычисляет открытый ключ $Y = [-x]G$, который передает проверяющему B.

Модифицированный протокол аутентификации Шнорра имеет следующий вид:

1. Абонент A генерирует случайное целое число k, $2 \leqslant k \leqslant n-2$, вычисляет точку эллиптической кривой $R = [k]G$ и отправляет проверяющему B $cert_A$ и значение R.

2. Проверяющий B генерирует случайное число a, которое передает абоненту A.

3. Абонент A вычисляет и передает проверяющему B значение $s = k + ax \pmod n$.

4. Если выполнено равенство $R = [s]G + [a]Y$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Полнота. Доказывающий A знает значение секрета x, поэтому он в состоянии ответить на оба вопроса. Проверяющий B убеждается в справедливости соотношений:

$$[s]G + [a]Y = [k + ax]G + [-ax]G = [k]G = R.$$

Корректность. Противник не знает секретного ключа x, поэтому для случая больших чисел a от B вероятность обмануть проверяющего B не превосходит 2^{-t}. Рассмотрим случай, когда $a \in \{0, 1\}$. Противник может подготовиться к запросу $a = 0$. В этом случае он сгенерирует случайное целое число k, $2 \leqslant k \leqslant q - 2$, вычислит $R = [k]G$ и отправит проверяющему B значение R. Тогда на запрос $a = 0$ противник передаст проверяющему B значение $s = k \pmod n$. В этом случае он не сможет ответить правильно на запрос $a = 1$. Так же противник может подготовиться к запросу $a = 1$. В этом случае он сгенерирует случайное число s, $1 \leqslant s \leqslant n - 1$, вычислит $R = [s]G + Y$. На первом шаге противник передаст значение R, а на запрос $a = 1$ передаст s. Поэтому он верно ответит на запрос $a = 1$, но не сможет ответить на запрос $a = 0$, так как придется решить трудную задачу дискретного логарифмирования.

Нулевое разглашение. Проверяющий B может самостоятельно получить любую тройку сообщений протокола (R, a, s). Для этого он может генерировать значения s, a, $1 \leqslant s \leqslant n - 1$, и вычислять $R = [s]G + [a]Y$. Такие тройки являются корректными тройками протокола и получены без участия абонента A. При этом множество всех корректных троек протокола, получаемых таким образом проверяющим B, совпадают с множеством всех допустимых троек протокола. Поэтому при наличии доказательства знания секрета x от абонента A в виде таких троек, которые проверяющий B может получить и сам, он не может получить из них дополнительную информацию о секрете x участника A.

Модификация трехпроходного протокола Шнорра на эллиптических кривых.

Абонент A выбирает случайные (секретные) числа x_i, $1 \leqslant x_1 \leqslant q-2$, и вычисляет значения открытых ключей $Y_i = [-x_i]G$, $i = 1, \ldots, m$.

На эллиптической кривой трехпроходный протокол аутентификации Шнорра примет следующий вид:

1. Доказывающий A генерирует случайное целое число k, где

$1 \leqslant k \leqslant q-2$, вычисляет $R = [k]G$ и отправляет проверяющему B точку эллиптической кривой R.

2. Проверяющий B генерирует случайную битовую строку $(a_1, \ldots, a_m) \in \{0,1\}^m$.

3. Абонент A вычисляет и передает проверяющему B значение:
$$s = k + \sum_{i=1}^{m} a_i x_i \pmod{q}.$$

4. Теперь если выполнено равенство:
$$R = [s]G + \sum_{i=1}^{m} [a_i]Y_i,$$
то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Модификация протокола Окамото на эллиптических кривых.

Пусть — эллиптическая кривая, известная участникам информационного процесса, G_1, G_2 — предварительно согласованные и опубликованные точки порядка n этой кривой. Абонент выбирает секретные числа x_1, x_2, $1 < x_1 < n$, $1 < x_2 < n$, и вычисляет открытый ключ $Y = [-x_1]G_1 + [-x_2]G_2$, который передает проверяющему B.

Модифицированный протокол аутентификации Окамото имеет следующий вид:

1. Абонент A генерирует случайные целые числа k_1, k_2, $2 \leqslant k_1 \leqslant n-2$, $2 \leqslant k_2 \leqslant n-2$, вычисляет $R = [k_1]G_1 + [k_2]G_2$ и отправляет проверяющему B $cert_A$ и значение R.

2. Проверяющий B генерирует случайное число a, которое передает абоненту A.

3. Абонент A вычисляет и передает проверяющему B значения:
$$\begin{cases} s_1 = k_1 + ax_1 \pmod{n}, \\ s_2 = k_2 + ax_2 \pmod{n}. \end{cases}$$

4. Если выполнено равенство $R = [s_1]G_1 + [s_2]G_2 + [a]Y$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Полнота. Доказывающий A знает значение секрета x, поэтому он в состоянии ответить на оба вопроса. Проверяющий B убеждается в справедливости соотношений:

$$[s_1]G_1 + [s_2]G_2 + [a]Y =$$

$$= [k_1 + ax_1]G_1 + [k_2 + ax_2]G_2 + [-ax_1]G_1 + [-ax_2]G_2 = R.$$

14.3.9. Протокол аутентификации Гиллоу-Куискатр (GQ)

Данный протокол основан на схеме RSA. В данном протоколе выбирают большие простые числа p и q, вычисляют $n = pq$, $\varphi(n) = (p-1)(q-1)$, выбирают элемент e, $2 < e < \varphi(n)$, $(e, \varphi(n)) = 1$, где φ — функция Эйлера. Числа e и n являются открытыми параметрами.

Абонент A генерирует случайный секретный ключ x, причем $1 < x < n$, $(x, n) = 1$, и вычисляет значение отрытого ключа $y = (x^{-1})^e \pmod n$.

Протокол аутентификации GQ имеет следующий вид:

1. Абонент A генерирует случайное целое число k, для которого выполнено $1 \leqslant k \leqslant n-2$, вычисляет $r = k^e \pmod n$ и отправляет проверяющему B $cert_A$ и значение r.

2. Проверяющий B генерирует случайное число a $(0 \leqslant a < e)$, которое передает абоненту A.

3. Абонент A вычисляет и передает проверяющему B значение $s = kx^a \pmod n$.

4. Если выполнено равенство $r = y^a s^e \pmod n$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Полнота протокола вытекает из соотношений

$$y^a s^e \equiv x^{-ea} k^e x^{ae} \equiv k^e \equiv r \pmod{n}.$$

Корректность. Рассмотрим случай, когда $a \in \{0, 1\}$. Противник может подготовиться к запросу $a = 0$. В этом случае он сгенерирует случайное целое число k, $1 \leqslant k \leqslant n-2$, вычислит $r = k^e \pmod{n}$ и отправит проверяющему B значение r. Тогда на запрос $a = 0$ противник передаст проверяющему B значение $s = k \pmod{p}$. В этом случае он не сможет ответить правильно на запрос $a = 1$. Так же противник может подготовиться к запросу $a = 1$. В этом случае он сгенерирует случайное число s, $1 \leqslant s \leqslant n-1$, вычислит $r = s^e y \pmod{n}$. На первом шаге протокола противник передаст значение r, а на запрос $a = 1$ передаст s. Поэтому он верно ответит на запрос $a = 1$, но не сможет ответить на запрос $a = 0$. Таким образом, вероятность обмана при t-кратной итерации не превосходит значения 2^{-t}.

Нулевое разглашение. Проверяющий B может самостоятельно получить любую тройку сообщений протокола (r, a, s). Для этого он может генерировать значения s, a, $1 \leqslant s \leqslant n-1$, $1 \leqslant a \leqslant n-1$, и вычислять $r = y^a s^e \pmod{n}$. Такие тройки являются корректными тройками протокола и получены без участия абонента A. При этом множество всех корректных троек протокола, получаемых таким образом проверяющим B, совпадают с множеством всех допустимых троек протокола. Поэтому при наличии доказательства знания секрета x от абонента A в виде таких троек, которые проверяющий B может получить и сам, он не может получить из них дополнительную информацию о секрете x участника A.

14.3.10. Протокол аутентификации на основе задачи о доказательстве изоморфизма графов

Пусть $G = (V, E_1)$ и $H = (V, E_2)$ — два графа с множеством вершин $V = \{1, \ldots, n\}$. Напомним, что графы G и H называются *изоморфными*, если существует такое биективное преобразование $\sigma \in S_n$ множества вершин V, при котором $(u, v) \in E_1$

тогда и только тогда, когда $(\sigma(u), \sigma(v)) \in E_2$ (будем обозначать $\sigma(G) = H$). Задача доказательства изоморфизма графов является NP-полной.

Доверенный центр T генерирует граф $G = (V, E_1)$, генерирует перестановку $\sigma \in S_n$, получает $\sigma(G) = H$ и передает G, H и σ абоненту A, уничтожая при этом у себя перестановку σ.

Протокол доказательства знания изоморфизма графов G и H состоит из следующих шагов:

1. Абонент A генерирует некоторую случайную перестановку $\tau \in S_n$ и отправляет проверяющему B граф $F = \tau(H)$, который изоморфен и графу G, и графу H.

2. Проверяющий B выбирает случайное значение $a \in \{0, 1\}$ и передает a доказывающему A (тем самым B просит доказать, что либо F изоморфен G, либо F изоморфен H).

3. Если $a = 1$, то абонент A передает B перестановку τ; если $a = 0$ — $\tau \circ \sigma$.

4. B проверяет равенство $\tau(H) = F$ или $(\tau \circ \sigma)(G) = F$ в зависимости от переданного ранее a.

Данный протокол повторяется t раз.

Полнота протокола очевидна.

Корректность. Противник не знает изоморфизма σ, поэтому может подготовиться к ответу только на один из двух вопросов проверяющего B. Поэтому вероятность обмана при t-кратной итерации не превосходит значения 2^{-t}.

Нулевое разглашение. Проверяющий B может самостоятельно получить любую перестановку $\tau \in S_n$ с тем свойством, что $\tau(H) = F$ или $\tau(G) = F$.

14.3.11. Протокол аутентификации на основе задачи о раскраске графа

Рассмотрим задачу построения протокола «доказательство знания с нулевым разглашением». При этом будем учитывать,

что каждый из участников может вести «нечестную» игру и пытаться обмануть другого. В качестве сложной задачи, решение которой известно абоненту A (переданное доверенным центром T), рассмотрим задачу раскраски графа тремя красками. Данная задача является NP-полной (т.е. для ее решения неизвестны алгоритмы, существенно более быстрые, чем метод перебора; при этом время решения задачи растет экспоненциально с ростом объема исходных данных).

В задаче о раскраске графа рассматривается граф с множеством вершин V и множеством ребер E. Абонент A знает правильную раскраску этого графа тремя красками (красной (R), зеленой (G), синей (B)). Правильная раскраска — это такая, когда любые две вершины, соединенные ребром, окрашены разными цветами. Приведем пример (рис. 14.1).

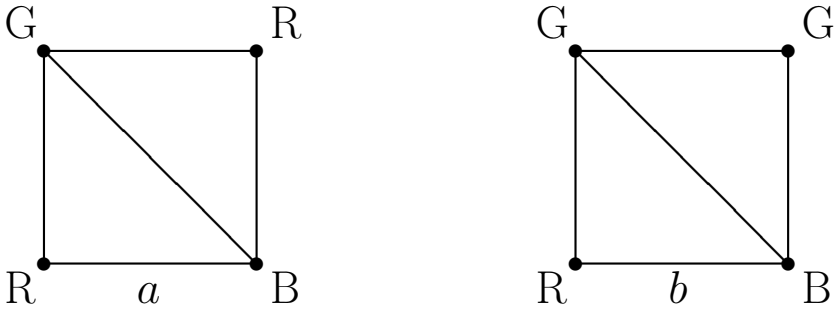


Рис. 14.1. Примеры раскрасок: a — правильная, b — неправильная

Для получения правильной раскраски графа тремя красками известны только экспоненциальные алгоритмы (у которых время решения задачи растет экспоненциально с ростом числа вершин и ребер в графе). Поэтому в случае больших $|V|$ и $|E|$ эта задача практически неразрешима.

Доверенный центр генерирует граф $H = (V, E)$, $V = \{1, 2, \ldots, k\}$, параллельно раскрашивая его правильной раскраской (это можно сделать за полиномиальное время). Абоненту A передается матрица смежности графа H и таблица правильной раскраски:

$$T: \begin{array}{|c|c|c|c|} \hline 1 & 2 & \ldots & k \\ \hline c(1) & c(2) & \ldots & c(k) \\ \hline \end{array},$$

где $c(i) \in \{R, G, B\}$, $i = 1, \ldots, k$.

Проверяющей стороне B передается только матрица смежности графа H.

Итак, пусть абонент A знает правильную раскраску графа H с большими $|V|$ и $|E|$. Этот абонент хочет доказать это абоненту B, но так, чтобы B ничего не узнал об этой раскраске.

Протокол доказательство состоит из множества одинаковых этапов. Опишем сначала один этап.

1. Абонент A выбирает случайно перестановку Π из трех букв R, G, B и перенумеровывает все вершины графа согласно этой перестановке. При этом понятно, что раскраска остается правильной. Например, если $\Pi = (G, B, R)$, то граф слева на рис. 14.1 превращается в граф на рис. 14.2.

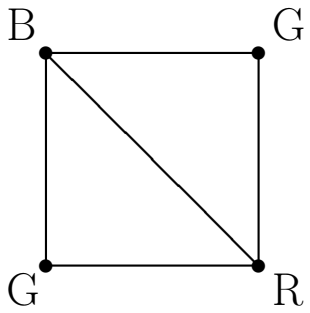


Рис. 14.2. Вариант раскраски после действия перестановки Π

Полученная таблица раскраски будет иметь такой вид:

$\widetilde{T}$:

1	2	...	k
$\Pi(c(1))$	$\Pi(c(2))$	...	$\Pi(c(k))$

Для каждой вершины $v \in V$ абонент A генерирует большое случайное число r и заменяет в нем два последних бита на:

00, если вершина v в таблице $\widetilde{T}$ окрашена в красный цвет,

01, если вершина v в таблице $\widetilde{T}$ окрашена в зеленый цвет,

10, если вершина v в таблице $\widetilde{T}$ окрашена в синий цвет.

Для каждой вершины $v \in V$ абонент A формирует данные, используемые в RSA, а именно, p_v, q_v, $n_v = p_v q_v$, e_v и d_v.

Абонент A вычисляет $m_v = r_v^{e_v} \pmod{n_v}$ и посылает абоненту B такую таблицу:

L :

1	2	...	k
n_1, e_1, m_1	n_2, e_2, m_2	...	n_k, e_k, m_k

.

2. Абонент B случайно выбирает ребро из множества E и сообщает абоненту A, какое именно ребро он выбрал.

3. В ответ абонент A высылает числа d_{v_1} и d_{v_2}, соответствующие вершинам этого ребра. После этого абонент B вычисляет:

$$\widetilde{m}_{v_1} = m_{v_1}^{d_{v_1}} \pmod{n_{v_1}} = r_{v_1}, \quad \widetilde{m}_{v_2} = m_{v_2}^{d_{v_2}} \pmod{n_{v_2}} = r_{v_2}$$

и сравнивает два младших бита в полученных числах. При правильной раскраске два младших бита в числах $\widetilde{m}_{v_1}$ и $\widetilde{m}_{v_2}$ должны быть различными. Если значения совпали, значит абонент A пытался обмануть абонента B, и на этом все заканчивается. Если не совпали, то весь описанный процесс повторяется $\alpha|E|$ раз, где $\alpha > 0$ — параметр.

Предложение 14.3. Если абонент A не располагает правильной раскраской, то вероятность того, что он может обмануть абонента B не превышает значения $e^{-\alpha}$. где e — основание натурального логарифма.

Доказательство. Пусть абонент A не располагает правильной раскраской. Тогда хотя бы для одного ребра из E вершины окрашены в один цвет. Вероятность того, что абонент B выберет это ребро равна $1/|E|$ (в этом случае абонент A будет разоблачен). Значит, вероятность того, что абонент A не разоблачен во время одного этапа протокола, не превышает $1-1/|E|$, а вероятность того, что он не будет разоблачен за $\alpha|E|$ этапов, не превышает $(1-1/|E|)^{\alpha|E|}$. Исходя из известного неравенства $1-x \leqslant e^{-x}$, получаем:

$$(1-1/|E|)^{\alpha|E|} \leqslant (e^{-1/|E|})^{\alpha|E|} = e^{-\alpha}. \qquad \square$$

14.3.12. Протокол аутентификации на основе задачи о нахождении гамильтонова цикла в графе

Блюм показал, что любое математическое утверждение может быть представлено в виде графа, причем доказательство этого утверждения соответствует гамильтонову циклу в этом

графе. Поэтому наличие протокола доказательства с нулевым разглашением для гамильтонова цикла означает, что доказательство любого математического утверждения может быть представлено в виде доказательства с нулевым разглашением.

Определение 14.1. *Гамильтоновым циклом* в графе называется непрерывный путь, проходящий через все вершины графа ровно по одному разу.

Например, на рис. 14.3 изображен граф G с гамильтоновым циклом (3, 4, 1, 2, 5).

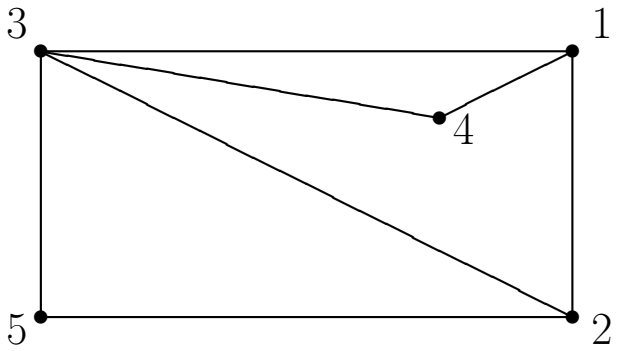


Рис. 14.3. Граф с гамильтоновым циклом (3, 4, 1, 2, 5)

Понятно, что если в графе n вершин (занумерованных числами $1, \ldots, n$) и в нем имеется гамильтонов цикл, то путем перебора всех перестановок симметрической группы S_n мы найдем гамильтонов цикл $(\tau(1), \ldots, \tau(n))$ для некоторой перестановки $\tau \in S_n$. Так как $|S_n| = n!$, то уже при сравнительно небольших значений n (например, $n = 100$) такой подход становится практически нереализуемым. Доказано, что задача нахождения гамильтонова цикла в графе является NP-полной (т.е. для ее решения неизвестны алгоритмы, существенно более быстрые, чем метод перебора).

Рассмотрим протокол, в котором абонент A будет доказывать абоненту B, что он знает гамильтонов цикл в некотором графе G так, чтобы абонент B не получил никаких знаний о самом этом цикле (доказательство с нулевым разглашением). Напомним, что «нулевое разглашением» означает, что независимо от числа реализаций протокола доказательства абонент B будет располагать точно такими же сведениями о гамильтоновом цик-

ле, какие он мог бы получить, просто изучая представленный ему граф G.

Пусть абонент A знает гамильтонов цикл в графе G из n вершин, который передал ему доверенный центр. Он может это доказывать абоненту B (и всем, кто имеет этот граф) с помощью описываемого ниже протокола.

Протокол доказательства состоит из следующих шагов:

1. Абонент A случайно выбирает перестановку $\sigma \in S_n$ и применяет ее к номерам вершин графа G, получив при этом граф $H = \sigma(G)$. Понятно, что графы G и H изоморфны. Зная гамильтонов цикл в графе G, абонент A знает гамильтонов цикл и в графе H. Граф H передается проверяющему B.

2. Абонент B, получив граф H, случайным образом выбирает $a \in \{0, 1\}$ и передает a абоненту A.

3. Если $a = 0$, то абонент A предоставляет абоненту B перестановку σ (тем самым показывая, что он знает изоморфизм графов G и H). Если $a = 1$, то абонент A предоставляет проверяющему B гамильтонов цикл графа H.

4. Проверяющий B проверяет, что в случае $a = 0$ предъявленная перестановка σ действительно переводит граф G в граф H, а в случае $a = 1$ проверяет гамильтонов цикл графа H.

Весть протокол повторяется t раз.

Абонент B задает два вопроса абоненту A по следующей причине. Если бы он задавал только первый вопрос, то абонент A, не зная в действительности гамильтонова цикла в графе G, мог бы предъявить абоненту B совсем другой граф с таким же количеством вершин и искусственно заложенным в него гамильтоновым циклом. Поэтому иногда абонент B иногда просит абонента A доказать изоморфизм графов G и H. При этом важно, что абонент A заранее не знает, какой из двух вопросов задаст абонент B.

Предложение 14.4. Вероятность обмана при t реализациях протокола не превосходит 2^{-t}.

Доказательство. Пусть некоторый абонент A не знает гамильтонов цикл в графе G. Тогда он может заранее подготовиться либо на первый вопрос (построив совсем другой граф с гамильтоновым путем) либо на второй вопрос. При этом если абонент A готовился к ответу на первый вопрос, а B задаст второй вопрос, то абонент A не сможет правильно на него ответить. И наоборот. Тогда обман вскроется. Поэтому вероятность успешности обмана равна вероятности угадывания номера вопроса. В предположении, что абонент B задает вопросы с одинаковой вероятностью, получаем, что вероятность обмана равна $1/2$. Вероятность же обмана при t реализациях протокола равна $1/2^t$. □

14.3.13. Протоколы аутентификации с нулевым разглашением знания на основе асимметричных шифров

Построение протоколов с нулевым разглашением можно реализовать, используя известные алгоритмы открытого шифрования. В качестве секретной информации, которой владеет доказывающая сторона A, будет использоваться секретный ключ x асимметричного шифра. Пусть D_x — алгоритм расшифрования на секретном ключе x, E_y — алгоритм шифрования на открытом ключе y. Проверяющая сторона шифрует некоторое сообщение M на открытом ключе y и передает криптограмму $C = E_y(M)$ абоненту A. Абонент A демонстрирует владение секретной информацией x тем, что расшифровывает сообщение своим секретным ключом: $M = D_x(C)$ и передает сообщение M проверяющей стороне B. Для проверяющей стороны B это не несет никакой дополнительной информации о секретном ключе x, так как у B до этого было то же самое сообщение M. В то же время, противник может выбрать произвольное значение и объявить его криптограммой, полученной в результате шифрования сообщения M, и попросить абонента A выполнить

процедуру расшифрования криптограммы. Если последний это сделает и раскроет восстановленное сообщение, то уже потенциально может иметь место утечка информации о секретном ключе x. Поэтому, при построении протоколов с нулевым разглашением знания, нужен некоторый механизм, который позволит владельцу секретного ключа (доказывающему A) до передачи восстановленного сообщения M проверяющему B убедиться в том, что последнее уже известно проверяющему B.

В качестве такого механизма могут использоваться алгоритмы хеширования (хеш-функции). Данный механизм используется в протоколах с нулевым разглашением, описанных в стандарте [60].

В стандарте [60] регламентируется формирование запроса в виде пары значений (C, H), где C — шифртекст, полученный путем шифрования некоторого сообщения M по открытому ключу доказывающего A и H — значение хеш-функции, вычисленное от сообщения M с использованием некоторой специфицированной хеш-функции $h : H = h(M)$. Получая запрос (C, H), доказывающий имеет возможность убедиться в том, что восстановленное им из шифртекста C сообщение M известно проверяющему. Для этого достаточно вычислить значение хеш-функции от восстановленного сообщения и сравнить его со значением второго элемента запроса.

В соответствии с [60] двухшаговый протокол с нулевым разглашением знания включает следующие шаги:

1. Проверяющий B выбирает произвольное сообщение M и, используя специфицированный алгоритм открытого шифрования E_y и открытый ключ y доказывающего, зашифровывает сообщение $: C = E_y(M)$. Затем, используя специфицированную хеш-функцию h, вычисляет значение хеш-функции от $: H = h(M)$. После этого он отправляет доказывающему пару значений (C, H) в качестве своего запроса.

2. Доказывающий A расшифровывает криптограмму , используя свой личный секретный ключ x, в результате чего получает сообщение$\tilde{}= D_x(C)$. Затем он вычисляет значение

хеш-функции от$\widetilde{\ }$: $\widetilde{H} = h(\widetilde{M})$, сравнивает значения $\widetilde{H}$ и H и, если $\widetilde{H} = H$, то отправляет проверяющему значение $\widetilde{M}$ в качестве своего ответа.

3. Если выполнено равенство $\widetilde{M} = M$, то проверяющий B принимает доказательство; если равенство не выполнено, то — отвергает.

Протокол аутентификации на основе шифра RSA

Доказывающий A выбирает два больших простых числа p и q, вычисляет $n = p \cdot q$, $\varphi(n) = (p-1) \cdot (q-1)$, где $\varphi(n)$ — функция Эйлера. Затем выбирается некоторое число e, $e < \varphi(n)$, взаимно простое с $\varphi(n)$, и находится решение сравнения $ed \equiv 1 \pmod{\varphi(n)}$. d — секретный ключ абонента A, пара n и e — открытый ключ.

Протокол аутентификации имеет следующий вид:

1. Проверяющий B генерирует случайное сообщение M, $0 < M < n$, вычисляет $C = M^e \pmod{n}$, $H = h(M)$ и отправляет доказывающему A пару значений (C, H).

2. Доказывающий A вычисляет $\widetilde{\ } = C^d \pmod{n}$, $\widetilde{H} = h(\widetilde{M})$. Если $\widetilde{H} = H$, то отправляет проверяющему значение $\widetilde{M}$ в качестве своего ответа.

3. После этого проверяющий B проверяет равенство $\widetilde{M} = M$.

Протокол аутентификации на основе шифра Эль-Гамаля и его модификация на эллиптических кривых

Абонент A выбирает некоторое большое простое число p, первообразный корень g по модулю p, достаточно большое число x (секретный ключ), $0 < x < p$, вычисляет значение открытого ключа $y = g^x \pmod{p}$.

Протокол аутентификации имеет следующий вид:

1. Проверяющий B генерирует случайное сообщение M, $0 < M < p$, генерирует случайным образом некоторое число k, $0 < k < p - 1$, вычисляет числа:

$$c_1 \equiv g^k \pmod{p},$$

$$c_2 \equiv M \cdot y^k \pmod{p},$$

вычисляет $H = h(M)$ и отправляет доказывающему A тройку значений $(_1, _2, H)$.

2. Доказывающий A вычисляет:

$$\widetilde{M} \equiv c_2 \cdot c_1^{p-1-x} \pmod{p}, \quad \widetilde{H} = h(\widetilde{M}).$$

Если $\widetilde{H} = H$, то отправляет проверяющему значение $\widetilde{M}$ в качестве своего ответа.

3. После этого проверяющий B проверяет равенство $\widetilde{M} = M$.

Приведем модификацию предыдущего протокола на эллиптических кривых.

Пусть q — некоторый (достаточно большой) простой делитель числа $|E|$, где E — эллиптическая кривая, и некоторая точка $G \in E$ имеет порядок q.

Общедоступные параметры системы: q, G, E. Абонент A выбирает секретный ключ x, $0 < x < q$, и вычисляет открытый ключ $Y = [x]G$.

Модифицированный протокол аутентификации имеет следующий вид:

1. Проверяющий B генерирует случайную точку $M = [r]G \in E$ для некоторого случайного r, $0 < r < q$, генерирует случайным образом некоторое число k, $0 < k < q$, вычисляет точки эллиптической кривой:

$$C_1 = [k]G,$$

$$C_2 = M + [k]Y.$$

вычисляет $H = h(M)$ и отправляет доказывающему A тройку значений (C_1, C_2, H).

2. Доказывающий A вычисляет:

$$\widetilde{M} = C_2 + [q - x]C_1, \quad \widetilde{H} = h(\widetilde{M}).$$

Если $\widetilde{H} = H$, то отправляет проверяющему значение $\widetilde{M}$ в качестве своего ответа.

3. После этого проверяющий B проверяет равенство $\widetilde{M} = M$.

Протокол аутентификации на основе криптосистемы Диффи-Хеллмана и его модификация на эллиптических кривых

Абонент A выбирает достаточно большое простое число p, некоторое число g, которое является первообразным корнем по модулю p, значение секретного ключа x, $1 < x < p - 1$, вычисляет значение открытого ключа $y = g^x \pmod p$.

Протокол аутентификации имеет следующий вид:

1. Проверяющий B генерирует случайное число k, $1 < k < p - 1$, вычисляет значения:

$$C = g^k \pmod p, \quad Z = y^k \pmod p, \quad H = h(Z)$$

и отправляет доказывающему A пару значений (C, H).

2. Доказывающий A вычисляет:

$$\widetilde{Z} = C^x \pmod p, \quad \widetilde{H} = h(\widetilde{Z}).$$

Если $\widetilde{H} = H$, то отправляет проверяющему значение $\widetilde{Z}$ в качестве своего ответа.

3. После этого проверяющий B проверяет равенство $\widetilde{Z} = Z$.

Приведем модификацию предыдущего протокола с использованием эллиптических кривых.

Пусть, как и ранее, q — некоторый (достаточно большой) простой делитель числа $|E|$, где E — эллиптическая кривая. Пусть некоторая точка $G \in E$ имеет порядок q. Абонент A выбирает случайное число x, $1 \leqslant x \leqslant q - 2$, и вычисляет значение $Y = [x]G$.

1. Проверяющий B генерирует случайное число k, $1 < k < q - 1$, вычисляет точки эллиптической кривой E и соответствующее значение хеш-функции:

$$= [k]G, \quad Z = [k]Y, \quad H = h(Z)$$

и отправляет доказывающему A пару значений (C, H).

2. Доказывающий A вычисляет:

$$\widetilde{Z} = [x]C, \quad \widetilde{H} = h(\widetilde{Z}).$$

Если $\widetilde{H} = H$, то отправляет проверяющему точку эллиптической кривой $\widetilde{Z}$ в качестве своего ответа.

3. После этого проверяющий B проверяет равенство $\widetilde{Z} = Z$.

Глава 15. Протоколы с нулевым разглашением

15.1. Протоколы привязки к биту

Протокол привязки к биту — примитивный криптографический протокол с двумя участниками (отправителем и получателем), посредством которого отправитель передает получателю бит информации (битовое обязательство) таким образом, что выполняются следующие два условия:

1. Связывание — после передачи бита получателю отправитель уже не может изменить его значение.
2. Сокрытие — получатель не может самостоятельно определить значение бита и узнает его только после выполнения отправителем так называемого этапа раскрытия.

Формально протокол привязки к биту можно записать следующим образом. Пусть $a \in \{0,1\}$ — обязательство, $f : \{0,1\} \times K \to R$ — некоторая функция, тогда сообщения протокола имеют вид:

$A \to B : r = f(a,k)$ (привязка);

$A \to B : a,\ k$ (раскрытие).

Участник A выбирает случайный элемент k, вычисляет $r = f(a,k)$ и отправляет участнику B. В дальнейшем для проверки обязательства участник открывает значение k, и участник B убеждается в том, что элемент a выбран на первом шаге.

Привязка к биту на основе хеш-функции. Приведем пример привязки к биту из [46]. Пусть h — бесключевая хеш-функция. Полагаем $f(a,k) = h(a \,||\, k)$, где $||$ — конкатенация. В данном случае выполнение первого свойства обеспечивается

тем, что хеш-функция h обладает свойством сложности подбора коллизий и второго прообраза, а второго — свойством односторонности.

Привязка к биту на основе протокола Шнорра. Пусть p — простое число, q — простой делитель числа $p-1$, $g \in \mathbb{Z}_p$, имеющий порядок q.

1. Абонент B выбирает случайное число x, для которого выполнено $1 \leqslant x \leqslant q-1$, и вычисляет значение $y = g^x \pmod p$, которое передается абоненту A.

2. Абонент A генерирует случайное число k, $1 \leqslant x \leqslant q-1$, вычисляет $r = g^k y^a \pmod p$ и передает r абоненту B.

3. Абонент A передает a и k абоненту B, который проверяет, выполняется ли сравнение $g^k y^a \equiv r \pmod p$.

Абонент B после второго шага не может извлечь никакой информации о бите a, так как k выбирается случайным образом и значение r в обоих случаях (при $a = 0$ и $a = 1$) является случайным элементом группы, порожденной g, и поэтому не несет никакой информации о значении бита a.

С другой стороны, абонент A может в этой ситуации обманывать, только если он умеет вычислять дискретные логарифмы по основанию g. Действительно, предположим, что абонент A может подобрать такие $k_1, k_2 \in \mathbb{Z}_q^*$, для которых $g^{k_1} y^0 = g^{k_2} y^1 \pmod p$. Тогда $k_1 = k_2 + x \pmod q$, $x = k_1 - k_2 \pmod q$.

Привязка к биту на основе протокола Шнорра с использованием эллиптических кривых. Пусть q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$, где $E_p(a,b)$ — эллиптическая кривая над полем $\mathbf{Z}_p$. Пусть некоторая точка $G \in E_p(a,b)$ имеет порядок q.

1. Абонент B выбирает случайное число x, для которого выполнено $1 \leqslant x \leqslant q-1$, и вычисляет точку эллиптической кривой $Y = [x]G$, которая передается абоненту A.

2. Абонент A генерирует случайное число k, $1 \leqslant x \leqslant q-1$, вычисляет $R = [k]G + [a]Y$ и передает R абоненту B.

3. Абонент A передает a и k абоненту B, который проверяет, выполняется ли равенство $[k]G + [a]Y = R$.

15.2. Применение протоколов привязки к биту в протоколах аутентификации с нулевым разглашением

Протокол привязки к биту — один из основных типов примитивных криптографических протоколов. Он находит многочисленные применения в криптографии. В работе [9] рассмотрен способ повышения эффективности доказательств с нулевым разглашением на примере протокола на основе задачи изоморфизма графов. Здесь мы приведем аналог этого метода для протокола аутентификации с нулевым разглашением на основе задачи поиска гамильтонова цикла в графе (см. параграф 14.3.12).

В этом протоколе главная проблема — большое количество раундов, растущее пропорционально размеру графа. Достаточно естественная идея — выполнить все эти последовательные раунды параллельно. На первом шаге A выбирает m случайных перестановок $\sigma_1,\ldots,\sigma_m$, вычисляет $H_1 = \sigma_1(G),\ldots,H_m = \sigma_m(G)$ и посылает все эти m графов проверяющей стороне B. На втором шаге B выбирает m случайных битов $a_1,\ldots,a_m$ и посылает их A, а на третьем A формирует все m требуемых перестановок и посылает их B.

Но будет ли такой протокол доказательством с нулевым разглашением? Прежде всего заметим, что этот протокол трехпроходный, а как показывает результат [57], трехпроходных доказательств с нулевым разглашением скорее всего не существует. Проверяющая сторона B формирует свои запросы $a_1,\ldots,a_m$, уже получив от A все графы $H_1,\ldots,H_m$, и может выбирать их (запросы) зависящими достаточно сложным образом от всех этих графов.

Зависимость $a_1,\ldots,a_m$ от $H_1,\ldots,H_m$ можно предотвратить следующим образом. Проверяющий B выбирает свои запросы в самом начале выполнения протокола, еще до того, как увидит

$H_1,\ldots, H_m$. Каждый бит a_i упаковывается в значение (блоб) r_i, и B посылает все (блобы) $r_1,\ldots, r_m$ доказывающему A. Только после этого A посылает B все графы $H_1,\ldots, H_m$. В ответ B открывает $a_1,\ldots, a_m$, а доказывающий A, получив $a_1,\ldots, a_m$, формирует требуемые перестановки и посылает их B.

До реализации протокола абонент A выбирает случайное число x, $1 \leqslant x \leqslant q-1$, которое держится в секрете и вычисляет значение открытого ключа $y = g^x \pmod p$, которое размещается в открытом справочнике или передается проверяющей стороне B.

1. Абонент B генерирует случайные числа k_i, $1 \leqslant k_i \leqslant q-1$, битовую строку $(a_1,\ldots,a_m) \in \{0,1\}^m$, вычисляет $r_i = g^{k_i} y^{a_i} \pmod p$, $i = 1,\ldots,m$, и передает набор $r_1,\ldots, r_m$ абоненту A.

2. Абонент A случайно выбирает перестановки $\sigma_i \in S_n$ и применяет их к номерам вершин графа G, получив при этом графы $H_i = \sigma_i(G)$, $i = 1,\ldots,m$, которые передаются абоненту B.

3. Абонент B, получив графы $H_1,\ldots, H_m$, передает абоненту A значения $k_1,\ldots, k_m$, $a_1,\ldots, a_m$.

4. При $a_i = 0$ абонент A фиксирует перестановку σ_i, при $a_i = 1$ — перестановку, являющуюся гамильтоновом циклом графа H_i, $i = 1,\ldots,m$. Данные перестановки передаются абоненту B.

5. Проверяющий B проверяет, что в случае $a_i = 0$ предъявленная перестановка σ_i действительно переводит граф G в граф H_i, а в случае $a_i = 1$ проверяет гамильтонов цикл графа H_i, $i = 1,\ldots,m$.

Также приведем модификацию данного протокола на эллиптических кривых. Пусть, как и ранее, q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$, где $E_p(a,b)$ — эллиптическая кривая над полем $\mathbf{Z}_p$. Пусть некоторая точка $G \in E_p(a,b)$ имеет порядок q. Абонент A выбирает случайное

число x, $1 \leqslant x \leqslant q-1$, и вычисляет значение открытого ключа $Y = [x]G$.

1. Абонент B генерирует случайные числа k_i, $1 \leqslant k_i \leqslant q-1$, битовую строку $(a_1, \ldots, a_m) \in \{0,1\}^m$, вычисляет точки эллиптической кривой $R_i = [k_i]G + [a_i]Y$, $i = 1, \ldots, m$, и передает набор $R_1, \ldots, R_m$ абоненту A.

2. Абонент A случайно выбирает перестановки $\sigma_i \in S_n$ и применяет их к номерам вершин графа G, получив при этом графы $H_i = \sigma_i(G)$, $i = 1, \ldots, m$, которые передаются абоненту B.

3. Абонент B, получив графы $H_1, \ldots, H_m$, передает абоненту A значения $k_1, \ldots, k_m$, $a_1, \ldots, a_m$.

4. При $a_i = 0$ абонент A фиксирует перестановку σ_i, при $a_i = 1$ — перестановку, являющуюся гамильтоновом циклом графа H_i, $i = 1, \ldots, m$. Данные перестановки передаются абоненту B.

5. Проверяющий B проверяет, что в случае $a_i = 0$ предъявленная перестановка σ_i действительно переводит граф G в граф H_i, а в случае $a_i = 1$ проверяет гамильтонов цикл графа H_i, $i = 1, \ldots, m$.

15.3. Протоколы типа «подбрасывание монеты по телефону»

Предположим, что двум участникам A и B необходимо бросить жребий. Если они оба физически находятся в одном и том же месте, задачу можно решить с помощью обычной процедуры подбрасывания монеты. Если же они удалены друг от друга и могут общаться лишь по каналу связи, то задача о жребии решается следующим образом [53].

Пусть p — простое число, q — простой делитель числа $p-1$, $g \in \mathbb{Z}_p$, имеющий порядок q.

1. Абонент B выбирает случайное число x, для которого выполнено $1 \leqslant x \leqslant q-1$, и вычисляет значение $y = g^x \pmod p$, которое передается абоненту A.

2. Абонент A генерирует случайное число k, $1 \leqslant x \leqslant q-1$, $a \in \{0,1\}$, вычисляет $r = g^k y^a \pmod p$ и передает r абоненту B.

3. Абонент B выбирает случайное значение $b \in \{0,1\}$ и передает b абоненту A.

4. Абонент A передает a и k абоненту B, который проверяет, выполняется ли сравнение $g^k y^a \equiv r \pmod p$. Если да, то результатом протокола будет $a \oplus b$.

В данном протоколе если хотя бы один участник выбрал свой бит случайным образом, то и результат $a \oplus b$ будет случайным.

Нетрудно видеть, что данный протокол легко модифицируется с использованием эллиптических кривых.

15.4. Аргумент с нулевым разглашением

Трехшаговое интерактивное доказательство с нулевым разглашением:

1) $A \to B:\ r$ (заявка);
2) $A \leftarrow B:\ a$ (запрос);
3) $A \to B:\ s$ (ответ)

можно переделать в неинтерактивный (автономный) одношаговый протокол, называемый *аргументом с нулевым разглашением*, полагая $a = h(r) : A \to B:\ (r, a, s)$. Проверка правильности доказательства проверяющим B проходит аналогичным образом.

При этом для данного протокола уже не выполняется свойство нулевого разглашения, но в теоретико-сложностном смысле такое доказательство не позволяет участнику B извлечь информацию о секрете участника A.

Приведем пример аргумента с нулевым разглашением из [46], который понадобится в следующем параграфе. В данном прото-

коле участник A выбирает свой голос $b \in \{-1, 1\}$, который должен оставаться неизвестным, а затем формирует доказательство, подтверждающее его выбор. Сначала рассмотрим интерактивный протокол доказательства:

1) $A \to B: \ r = (r_0, r_1, r_2)$;

2) $A \leftarrow B: \ a$;

3) $A \to B: \ s = (d_1, d_2, s_1, s_2)$.

Пусть p — достаточно большое простое число, q — большой простой делитель числа $p-1$, g — элемент мультипликативной группы $\mathbb{Z}_p^*$, имеющий порядок q, $h = g^u \pmod p$ для некоторого u, $0 < u < q$. Участник A публикует значение свидетельства:

$$r_0 = g^k h^b \pmod p, \quad 0 < k < q, \quad b \in \{-1, 1\},$$

скрывающего поданный им голос (битовое обязательство) $b \in \{-1, 1\}$. Потом выбирает случайные элементы $0 < d, z, w < q$, вычисляет:

$$r_1 = \begin{cases} g^z (r_0 h)^{-d} \pmod p, & b = 1, \\ g^w \pmod p, & b = -1, \end{cases}$$

$$r_2 = \begin{cases} g^w \pmod p, & b = 1, \\ g^z (r_0 h^{-1})^{-d} \pmod p, & b = -1 \end{cases}$$

и публикует их (отправляет участнику B). Затем после получения запроса a от участника B отвечает четверкой:

$$(d_1, d_2, s_1, s_2) = \begin{cases} (d, \widetilde{d}, z, \widetilde{z}), & b = 1, \\ (\widetilde{d}, d, \widetilde{z}, z), & b = -1, \end{cases}$$

где $\widetilde{d} = a - d \pmod q$, $\widetilde{z} = w + k\widetilde{d} \pmod q$.

Участник B проверяет равенства:

$$a = d_1 + d_2 \pmod q,$$

$$g^{s_1} = r_1 (r_0 h)^{d_1} \pmod p,$$

$$g^{s_2} = r_2 (r_0 h^{-1})^{d_2} \pmod p.$$

Чтобы получить из данного протокола аргумент с нулевым разглашением, полагаем $a = h(r_0, r_1, r_2) \pmod q$, где h — хеш-

функция. В автономной версии протокола передается одно сообщение:

$$A \to B:\ (r_0, r_1, r_2),\ h(r_0, r_1, r_2),\ (d_1, d_2, s_1, s_2).$$

Модификация данного протокола на эллиптических кривых

Пусть E — эллиптическая кривая над некоторым конечным полем F, q — некоторый достаточно большой простой делитель числа $|E|$, G — некоторая точка эллиптической кривой, имеющая порядок q, $H = [u]G$ для некоторого $0 < u < q$.

Участник A публикует значение свидетельства:

$$R_0 = [k]G + [b]H, \quad 0 < k < q, \quad b \in \{-1, 1\},$$

скрывающего поданный им голос (битовое обязательство) $b \in \{-1, 1\}$. Потом выбирает случайные элементы $0 < d, z, w < q$, вычисляет:

$$R_1 = \begin{cases} [z]G + [-d](R_0 + H), & b = 1, \\ [w]G, & b = -1, \end{cases}$$

$$R_2 = \begin{cases} [w]G, & b = 1, \\ [z]G + [-d](R_0 - H), & b = -1 \end{cases}$$

и публикует их (отправляет участнику B). Затем после получения запроса a от участника B отвечает четверкой:

$$(d_1, d_2, s_1, s_2) = \begin{cases} (d, \widetilde{d}, z, \widetilde{z}), & b = 1, \\ (\widetilde{d}, d, \widetilde{z}, z), & b = -1, \end{cases}$$

где $\widetilde{d} = a - d \pmod q$, $\widetilde{z} = w + k\widetilde{d} \pmod q$.

Участник B проверяет равенства:

$$a = d_1 + d_2 \pmod q,$$

$$[s_1]G = R_1 + [d_1](R_0 + H),$$

$$[s_2]G = R_2 + [d_2](R_0 - H).$$

15.5. Протоколы электронного голосования

Протокол голосования — прикладной криптографический протокол, позволяющий проводить процедуру голосования, в которой избирательные бюллетени существуют только в электронной форме. Является криптографическим протоколом, т.к. обеспечивает тайный характер голосования. Основное свойство протокола голосования — универсальная проверяемость, т.е. предоставление возможности всякому желающему, включая сторонних наблюдателей, в любой момент времени проверить правильность подсчета голосов.

15.5.1. Протокол голосования на основе протокола Шаума-Педерсена

Рассмотрим протокол, приведенный в работе [9]. Пусть в голосовании участвуют n избирателей $P_1,\ldots,\ P_n$, которые являются абонентами некоторой сети и подают свои голоса в электронной форме: «за» и «против», которые соответственно представимы значениями 1 и -1. К протоколу предъявим два основных требования:

1) голосование должно быть тайным;

2) должна быть обеспечена правильность подсчета голосов.

Пусть T — центр подсчета голосов. Будем предполагать, что центр честный и пользуется безусловным доверием всех избирателей.

Пусть p — достаточно большое простое число, q — большой простой делитель числа $p-1$, g, h — элементы мультипликативной группы $\mathbb{Z}_p^*$, имеющие порядок q. Доверенный центр T выбирает секретный ключ x, $0 < x < q$, и публикует в открытом доступе открытый ключ $y = g^x \pmod p$. Каждый избиратель P_i посылает центру сообщение, содержащее идентификатор этого избирателя и его голос $a_i \in \{-1, 1\}$, зашифрованный с помощью вероятностного шифра на ключе y следующим образом: $u_i = g^{k_i} \pmod p$, $v_i = y^{k_i} h^{a_i} \pmod p$, (u_i, v_i) — бюллетень голосования избирателя P_i (передается центру T), где

k_i — некоторое случайное число, $0 < k_i < q$. Центр проверяет соответствие поданных бюллетеней спискам избирателей, расшифровывает бюллетени и отбрасывает недействительные (в которых голоса отличны от -1 и 1), подсчитывает и публикует итог. Расшифрование центром T бюллетеня (u_i, v_i) происходит следующим образом: вычисляется $v_i u_i^{q-x} = h^{a_i} \pmod p$; так как $a_i \in \{-1, 1\}$, то из h^{a_i} легко находится a_i. Недостатком этого метода является тот факт, что анонимность голосования обеспечивается для всех, кроме центра T, которому известно, как проголосовал каждый участник.

Предположим, что для проведения голосования создано табло — хранилище информации, в котором для каждого избирателя выделена отдельная строка. Эта строка содержит, например, полные данные избирателя, и в эту строку он помещает свой бюллетень. Предполагается, что табло доступно на чтение всем участникам голосования, а также сторонним наблюдателям. По истечении срока подачи голосов табло «закрывается», т.е. фиксируется его состояние. После этого выделяется некоторое время, в течение которого каждый избиратель проверяет содержимое своей строки на табло. Все претензии разбираются, при необходимости вносятся соответствующие изменения и, когда все избиратели удовлетворены, содержимое табло фиксируется окончательно.

После этого центр T вычисляет $S = \sum_{i=1}^{n} a_i$ и публикует итог голосования S. Поскольку все бюллетени находятся на табло, любой избиратель, а также всякий сторонний наблюдатель, может вычислить:

$$A = \prod_{i=1}^{n} u_i = \prod_{i=1}^{n} g^{k_i} \pmod p, \quad B = \prod_{i=1}^{n} v_i = \prod_{i=1}^{n} y^{k_i} h^{a_i} \pmod p.$$

Обозначим $C = \prod_{i=1}^{n} y^{k_i} \pmod p$. Заметим, что $C = A^x \pmod p$. Если центр правильно подсчитал голоса, то должно выполняться равенство $h^S = \prod_{i=1}^{n} h^{a_i} \pmod p$. Поэтому, если B поделить на h^S, то должно получиться значение C. Пусть $\widetilde{C} = B \cdot h^{-S} \pmod p$. Проблема в том, что проверяющий не знает значения C и не может самостоятельно выяснить, верно

ли, что $C = \widetilde{C}$. Но нетрудно проверить, что должно выполняться сравнение $\widetilde{C} = A^x \pmod p$. Поэтому проверяющий может потребовать от центра доказательство следующего факта: дискретный логарифм $\widetilde{C}$ по основанию A равен дискретному логарифму y по основанию g. Приведем предназначенный для этой цели протокол Шаума и Педерсена [54, 55].

1. Доказывающий случайным образом выбирает k, $0 < k < q$, вычисляет $r_1 = g^k \pmod p$, $r_2 = A^k \pmod p$ и передает r_1, r_2 проверяющему.

2. Проверяющий генерирует случайное число a, $0 \leqslant a < q$, которое передает доказывающему.

3. Доказывающий вычисляет $s = k + ax \pmod q$ и передает s проверяющему.

4. Проверяющий убеждается, что $g^s = r_1 y^a \pmod p$ и $A^s = r_2 \widetilde{C}^a \pmod p$.

Таким образом, центр T может доказать утверждение $\widetilde{C} = A^x \pmod p$ каждому желающему.

Модификация протокола голосования на эллиптических кривых.

Пусть E — эллиптическая кривая над некоторым конечным полем F, q — некоторый достаточно большой простой делитель числа $|E|$, G, H — некоторые точки эллиптической кривой, имеющие порядок q. Доверенный центр T выбирает секретный ключ x, $0 < x < q$, и публикует в открытом доступе открытый ключ $Y = [x]G$.

Каждый избиратель P_i посылает центру T сообщение, содержащее идентификатор этого избирателя и его голос $a_i \in \{-1, 1\}$, зашифрованный с помощью вероятностного шифра на ключе Y следующим образом: $U_i = [k_i]G$, $V_i = [k_i]Y + [a_i]H$, (U_i, V_i) — бюллетень голосования (передается центру T), где k_i — некоторое случайное число, $0 < k_i < q$. Центр расшифровывает бюллетени, подсчитывает и публикует

итог. Расшифрование бюллетеня (U_i, V_i) происходит следующим образом: вычисляется $V_i + [q - x]U_i = [a_i]H$; так как $a_i \in \{-1, 1\}$, то из $[a_i]H$ легко находится a_i.

После этого центр T вычисляет $S = \sum_{i=1}^{n} a_i$ и публикует итог голосования S. Поскольку все бюллетени находятся в некотором хранилище данных, то любой избиратель, а также всякий сторонний наблюдатель, может вычислить:

$$A = \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} [k_i]G, \quad B = \sum_{i=1}^{n} V_i = \sum_{i=1}^{n} ([k_i]Y + [a_i]H).$$

Обозначим $C = \sum_{i=1}^{n} [k_i]Y$. При этом $C = [x]A$. Если центр правильно подсчитал голоса, то должно выполняться равенство $[S]H = \sum_{i=1}^{n} [a_i]H$. Поэтому, если из B вычесть $[S]H$, то должно получиться значение C. Пусть $\widetilde{C} = B - [S]H$. Проверяющий не знает значения C и не может самостоятельно выяснить, верно ли, что $C = \widetilde{C}$. Но при этом нетрудно проверить, что должно выполняться равенство $\widetilde{C} = [x]A$. Поэтому проверяющий может потребовать от центра доказательство следующего факта: в группе точек эллиптической кривой дискретный логарифм $\widetilde{C}$ по основанию A равен дискретному логарифму Y по основанию G.

Приведем модификацию протокола Шаума и Педерсена на эллиптических кривых:

1. Доказывающий случайным образом выбирает k, $0 < k < q$, вычисляет $R_1 = [k]G$, $R_2 = [k]A$ и передает R_1, R_2 проверяющему.

2. Проверяющий генерирует случайное число a, $0 \leqslant a < q$, которое передает доказывающему.

3. Доказывающий вычисляет $s = k + ax \pmod{q}$ и передает s проверяющему.

4. Проверяющий убеждается, что $[s]G = R_1 + [a]Y$ и $[s]A = R_2 + [a]\widetilde{C}$.

15.5.2. Протокол Крамера-Франклина-Шонмейкерса-Янга и его модификация на эллиптических кривых

Рассмотрим более сложный протокол электронного голосования. Задача ставится следующим образом. Пусть в голосовании участвуют n избирателей $P_1,\ldots,P_n$, которые являются абонентами некоторой сети и подают свои голоса в электронной форме: «за» и «против», которые соответственно представимы значениями 1 и -1. Имеется m счетных комиссий, которые создаются для обеспечения анонимности и предотвращения фальсификации итогов голосования. К протоколу предъявим следующие требования:

1) голосуют только уполномоченные избиратели;

2) любой участник имеет право отдать не более одного голоса;

3) ни один из участников не может знать, как проголосовал другой;

4) никто не может дублировать чужой голос;

5) конечный результат будет подсчитан корректно;

6) любой желающий может проверить правильность результата;

7) протокол должен работать и в те случаи, если некоторые участники ведут себя нечестно.

Рассмотрим протокол, предложенный в 1996 г. (A. Cramer, M. Franclin, B. Shoenmakers, M. Yung [56]). Сначала каждая комиссия фиксирует закрытый ключ и публикует открытый ключ.

Пусть p — достаточно большое простое число, q — большой простой делитель числа $p-1$, g — элемент мультипликативной группы $\mathbb{Z}_p^*$, имеющий порядок q, $h = g^u \pmod p$ для некоторого u, $0 < u < q$, причем нахождение $\log_g h$ должно являться трудной задачей.

1. Заполнение бюллетеня избирателями. Избиратель P_i выбирает голос $a_i \in \{-1, 1\}$ и случайный элемент $k_i \in \mathbb{Z}_q$. Затем он публикует свидетельство $r_{0i} = g^{k_i} h^{a_i}$, $i = 1,\ldots,n$. В результате в общем доступе будут свидетельства всех участников $r_{01},\ldots,r_{0n}$. Также избиратель P_i выполняет автономную вер-

сию протокола доказательства знания, рассмотренного в параграфе 15.4. Затем избиратель P_i публикует значения (r_0, r_1, r_2), a, (d_1, d_2, s_1, s_2).

2. Передача бюллетеней комиссиям. Для передачи бюллетеней с голосами избирателей счетным комиссиям используется совершенная проверяемая схема разделения секрета Педерсена-Шамира (см. параграф 13.1.3): i-й избиратель выбирает два многочлена над полем $\mathbb{Z}_q$ степени T, $0 < T < m$:

$$U_i(x) = k_i + k_{1i}x + \ldots + k_{Ti}x^T \in \mathbb{Z}_q[x],$$

$$V_i(x) = a_i + a_{1i}x + \ldots + a_{Ti}x^T \in \mathbb{Z}_q[x],$$

где коэффициенты k_{ji}, a_{ji} — случайные числа из $\mathbb{Z}_q$, $1 \leqslant j \leqslant T$. Значения $(x_j, y_{ij}, z_{ij}) = (x_j, U_i(x_j), V_i(x_j))$, $x_j \in \mathbb{Z}_q^*$ попарно различны, $j = 1, \ldots, m$, являются долями $(m, T+1)$ пороговой схемы разделения секрета (k_i, a_i). Здесь значение T определяется тем, что если не произошло никакого сговора более чем T избирательных комиссий, то невозможно вычислить, как голосовал отдельный участник. В то же время, выборы будут успешны, если, по крайней мере, $T+1$ избирательных комиссий действуют правильно.

Также для данных коэффициентов P_i вычисляет проверочные значения:

$$B_{i0} = g^{k_i}h^{a_i},\ B_{i1} = g^{k_{1i}}h^{a_{1i}}, \ldots,\ B_{iT} = g^{k_{Ti}}h^{a_{Ti}},$$

которые публикуются в открытом доступе. Избиратель P_i шифрует значения (x_j, y_{ij}, z_{ij}) на открытом ключе j-й избирательной комиссии, после чего ей передаются полученные значения, $j = 1, \ldots, m$. j-я комиссия после расшифрования и восстановления значений (x_j, y_{ij}, z_{ij}) делает проверку

$$g^{y_{ij}}h^{z_{ij}} \equiv B_{i0}\,(B_{i1})^{x_j} \ldots (B_{iT})^{x_j^T} \pmod p.$$

Итак, каждая счетная комиссия имеет следующие наборы:

$$\begin{array}{lccc} 1: & (x_1, y_{11}, z_{11}), & \ldots, & (x_1, y_{n1}, z_{n1}), \\ & & \ldots & \\ m: & (x_m, y_{1m}, z_{1m}), & \ldots, & (x_m, y_{nm}, z_{nm}). \end{array}$$

3. Подсчет голосов. Каждая j-я комиссия подсчитывает и публикует значения:

$$y_j = \sum_{i=1}^{n} y_{ij}, \quad z_j = \sum_{i=1}^{n} z_{ij}.$$

Теперь каждый желающий может проверить корректность опубликованных данных, проверив равенства:

$$\prod_{i=1}^{n} \left(r_{0i} \prod_{l=1}^{T} B_{il}^{x_j^l} \right) = g^{y_j} h^{z_j}, \quad j = 1, \ldots, m,$$

так как:

$$\prod_{i=1}^{n} \left(r_{0i} \prod_{l=1}^{T} B_{il}^{x_j^l} \right) = \prod_{i=1}^{n} \left(g^{k_i} h^{a_i} \prod_{l=1}^{T} \left(g^{k_{li}} h^{a_{li}} \right)^{x_j^l} \right) =$$

$$= \prod_{i=1}^{n} g^{k_i + k_{1i} x_j + \ldots + k_{Ti} x_j^T} h^{a_i + a_{1i} x_j + \ldots + a_{Ti} x_j^T} = \prod_{i=1}^{n} g^{U_i(x_j)} h^{V_i(x_j)} =$$

$$\prod_{i=1}^{n} g^{y_{ij}} h^{z_{ij}} = g^{\sum_{i=1}^{n} y_{ij}} h^{\sum_{i=1}^{n} z_{ij}} = g^{y_j} h^{z_j}.$$

Заметим, что для любого $j = 1, \ldots, m$ значение z_j является значением некоторого многочлена над полем $\mathbb{Z}_q$ степени не более T :

$$z_j = \sum_{i=1}^{n} z_{ij} = \sum_{i=1}^{n} V_i(x_j) =$$

$$= \left(\sum_{i=1}^{n} a_i \right) + \left(\sum_{i=1}^{n} a_{1i} \right) x_j + \ldots + \left(\sum_{i=1}^{n} a_{Ti} \right) x_j^T.$$

Поэтому для определения итога голосования $\sum_{i=1}^{n} a_i$ достаточно в множестве пар точек $\{(x_1, z_1), \ldots, (x_m, z_m)\}$ выделить любое $(T+1)$-элементное подмножество $\{(\widetilde{x}_0, \widetilde{z}_0), \ldots, (\widetilde{x}_T, \widetilde{z}_T)\}$ и вычислить

$$\sum_{i=0}^{T} \widetilde{z}_i \prod_{\substack{0 \leqslant j \leqslant T \\ j \neq i}} \frac{\widetilde{x}_j}{\widetilde{x}_j - \widetilde{x}_i} = \sum_{i=1}^{n} a_i.$$

Модификация протокола на эллиптических кривых

Пусть E — эллиптическая кривая над некоторым конечным полем F, q — некоторый достаточно большой простой делитель числа $|E|$, G — некоторая точка эллиптической кривой, имеющая порядок q, $H = [u]G$ для некоторого $0 < u < q$, причем нахождение значения u по известному H должно являться трудной задачей.

1. Заполнение бюллетеня избирателями. Избиратель P_i выбирает голос $a_i \in \{-1, 1\}$ и случайный элемент $k_i \in \mathbb{Z}_q$. Затем он публикует свидетельство:

$$R_{0i} = [k_i]G + [a_i]H, \quad i = 1, \ldots, n,$$

скрывающее поданный им голос (битовое обязательство). В результате в общем доступе будут свидетельства всех участников $R_{01}, \ldots, R_{0n}$. Также избиратель P_i выполняет автономную версию протокола доказательства знания, рассмотренного в параграфе 15.4. Затем избиратель P_i публикует значения $(R_0, R_1, R_2),\ a,\ (d_1, d_2, s_1, s_2)$.

2. Передача бюллетеней комиссиям. Для передачи бюллетеней с голосами избирателей счетным комиссиям используется совершенная проверяемая схема разделения секрета Педерсена-Шамира: i-й избиратель выбирает два многочлена над полем $\mathbb{Z}_q$ степени T, $0 < T < m$:

$$U_i(x) = k_i + k_{1i}x + \ldots + k_{Ti}x^T \in \mathbb{Z}_q[x],$$

$$V_i(x) = a_i + a_{1i}x + \ldots + a_{Ti}x^T \in \mathbb{Z}_q[x],$$

где коэффициенты k_{ji}, a_{ji} — случайные числа из $\mathbb{Z}_q$, $1 \leqslant j \leqslant T$. Значения $(x_j, y_{ij}, z_{ij}) = (x_j, U_i(x_j), V_i(x_j))$, $x_j \in \mathbb{Z}_q^*$ попарно различны, $j = 1, \ldots, m$, являются долями $(m, T+1)$ пороговой схемы разделения секрета (k_i, a_i).

Также для данных коэффициентов P_i вычисляет проверочные значения:

$$B_{i0} = [k_i]G + [a_i]H,\ B_{i1} = [k_{1i}]G + [a_{1i}]H, \ldots,\ B_{iT} = [k_{Ti}]G + [a_{Ti}]H,$$

которые публикуются в открытом доступе.

Избиратель P_i шифрует значения (x_j, y_{ij}, z_{ij}) на открытом ключе j-й избирательной комиссии, после чего ей передаются

полученные значения, $j = 1, \ldots, m$. j-я комиссия после расшифрования и восстановления значений (x_j, y_{ij}, z_{ij}) делает проверку:

$$[y_{ij}]G + [z_{ij}]H = B_{i0} + [x_j]B_{i1} + \ldots + [x_j^T]B_{iT}.$$

3. Подсчет голосов. Каждая j-я комиссия подсчитывает и публикует значения:

$$y_j = \sum_{i=1}^{n} y_{ij}, \quad z_j = \sum_{i=1}^{n} z_{ij}.$$

Теперь каждый желающий может проверить корректность опубликованных данных, проверив равенства:

$$\sum_{i=1}^{n} \left(R_{0i} + \sum_{l=1}^{T} [x_j^l] B_{il} \right) = [y_j]G + [z_j]H, \quad j = 1, \ldots, m,$$

так как:

$$\sum_{i=1}^{n} \left(R_{0i} + \sum_{l=1}^{T} [x_j^l] B_{il} \right) =$$

$$= \sum_{i=1}^{n} \left([k_i]G + [a_i]H + \sum_{l=1}^{T} [x_j^l] \left([k_{li}]G + [a_{li}]H \right) \right) =$$

$$= \sum_{i=1}^{n} ([k_i + k_{1i}x_j + \ldots + k_{Ti}x_j^T]G + [a_i + a_{1i}x_j + \ldots + a_{Ti}x_j^T]H) =$$

$$= \sum_{i=1}^{n} ([U_i(x_j)]G + [V_i(x_j)]H) =$$

$$= \sum_{i=1}^{n} ([y_{ij}]G + [z_{ij}]H) = [y_j]G + [z_j]H.$$

Для определения итога голосования $\sum_{i=1}^{n} a_i$ достаточно в множестве пар точек $\{(x_1, z_1), \ldots, (x_m, z_m)\}$ выделить любое $(T+1)$-элементное подмножество $\{(\widetilde{x}_0, \widetilde{z}_0), \ldots, (\widetilde{x}_T, \widetilde{z}_T)\}$ и вычислить:

$$\sum_{i=0}^{T} \widetilde{z}_i \prod_{\substack{0 \leqslant j \leqslant T \\ j \neq i}} \frac{\widetilde{x}_j}{\widetilde{x}_j - \widetilde{x}_i} = \sum_{i=1}^{n} a_i.$$

Глава 16. Протоколы передачи ключей

Одной из самых сложных задач управления ключами, возникающих в криптографии, является формирование общих секретных ключей участников криптосистем.

Протокол распределения ключей — протокол получения пользователями ключей, необходимых для функционирования криптографической системы. Различают следующие типы протоколов распределения ключей:

- протоколы передачи (уже сгенерированных) ключей;
- протоколы совместной выработки общего ключа (открытое распределение ключей);
- схемы предварительного распределения ключей.

16.1. Передача ключей с использованием симметричного шифрования

Протоколы распределения ключей, основанные на симметричных криптосистемах, разделяются на два больших класса: протоколы без доверенного центра (двусторонние протоколы) и протоколы с доверенным центром (трехсторонние протоколы).

16.1.1. Двусторонние протоколы

Рассмотрим двусторонние протоколы передачи ключей с использованием симметричного шифрования.

Пусть стороны A и B заранее обладают общей секретной информацией — секретный ключ k_{AB} для симметричного блочного шифра E. Тогда для передачи ключа k стороны могут использовать одностороннюю передачу:

$$A \to B: \ E_{k_{AB}}(k, t, B),$$

где t — метка времени, B — идентификатор участника B (для краткости идентификаторы обозначены теми же символами, что и сами участники). Проверяющий B расшифровывает данное сообщение и проверяет соответствие допустимому интервалу временной метки и совпадение полученного и собственного идентификаторов. Если не передавать метки времени, то противник может осуществить повторную передачу того же сообщения. Если же не указывать идентификатор адресата, то противник может отправить перехваченное сообщение обратно абоненту A.

К недостаткам этого протокола можно отнести то, что ключ целиком определяется только одним участником.

Если предъявить к протоколу требование проведения аутентификации сеанса в целях более надежной аутентификации источника, то можно использовать следующий протокол «запрос–ответ»:

$$\begin{aligned} &1)\ \ A \leftarrow B:\ r_B; \\ &2)\ \ A \rightarrow B:\ E_{k_{AB}}(k, r_B, B), \end{aligned}$$

где r_B — случайное число, сгенерированное абонентом B и переданное абоненту A в начале сеанса.

Если требуется двусторонняя аутентификация, то можно модифицировать последний протокол:

$$\begin{aligned} &1)\ \ A \leftarrow B:\ r_B; \\ &2)\ \ A \rightarrow B:\ E_{k_{AB}}(k, r_A, r_B, B); \\ &3)\ \ A \leftarrow B:\ E_k(r_A). \end{aligned}$$

В данном протоколе участник A может убедиться, что он имеет дело именно с участником B и что участник B получил правильное значение ключа k.

Исходный протокол можно модифицировать так, чтобы искомый ключ k генерировался не одной стороной, а являлся результатом двустороннего обмена. Пусть участники A и B помимо случайных чисел r_A и r_B генерируют случайные числа k_A и

k_B соответственно. Тогда в результате выполнения протокола:

$$\begin{aligned}
&1)\ A \leftarrow B:\ r_B;\\
&2)\ A \rightarrow B:\ E_{k_{AB}}(k_A, r_A, r_B, B);\\
&3)\ A \leftarrow B:\ E_{k_{AB}}(k_B, r_B, r_A, A)
\end{aligned}$$

каждая из сторон может вычислить общий ключ k с помощью некоторой функции f по правилу $k = f(k_A, k_B)$. При этом в данном протоколе ни одна из сторон не может заранее предсказать значение ключа k. Данный протокол построен на основе «протокола рукопожатия».

Протокол рукопожатия для процедуры удаленного вызова. Рассмотрим пример двустороннего протокола передачи ключа, предложенного в 1985 г.:

$$\begin{aligned}
&1)\ A \rightarrow B:\ A,\ E_{k_{AB}}(r_A);\\
&2)\ A \leftarrow B:\ E_{k_{AB}}(r_A + 1, r_B);\\
&3)\ A \rightarrow B:\ E_{k_{AB}}(r_B + 1);\\
&4)\ A \leftarrow B:\ E_{k_{AB}}(k, \widetilde{r}_B),
\end{aligned}$$

где $\widetilde{r}_B$ — случайное число для последующего сеанса и для усложнения криптоанализа значения ключа k. Здесь первые три сообщения реализуют взаимную аутентификацию сторон, четвертое — передачу ключа, причем эти две части протокола никак не связаны между собой, что позволяет нарушителю повторно передать последнее сообщение старого протокола в новый, поэтому абонент A будет использовать старое значение ключа из прошлого протокола. Поэтому был предложен исправленный вариант протокола:

$$\begin{aligned}
&1)\ A \rightarrow B:\ A,\ r_A;\\
&2)\ A \leftarrow B:\ E_{k_{AB}}(r_A + 1, k);\\
&3)\ A \rightarrow B:\ E_k(r_A + 1);\\
&4)\ A \leftarrow B:\ \widetilde{r}_B,
\end{aligned}$$

где $\widetilde{r}_B$ — случайное число для последующего сеанса. Данный протокол имеет уязвимости. Противник может выступать от абонента B, который при этом вообще не участвует в обмене.

Использование односторонней функции. В этом протоколе участник A самостоятельно генерирует новый сеансовый

ключ k :

$$\begin{aligned}
&1)\ \ A \leftarrow B:\ \ B,\ r_B;\\
&2)\ \ A \rightarrow B:\ \ A,\ E_{k_{AB}}(h(r_B), r_A, A, k);\\
&3)\ \ A \leftarrow B:\ \ B,\ E_k(h(r_A)).
\end{aligned}$$

Получив сообщение на втором шаге, участник B расшифровывает его и проверяет правильность полученного значения $h(r_B)$. Затем он вычисляет проверочное значение $h(r_A)$ и отправляет участнику A, зашифровав на новом ключе k. Теперь участник A проверяет правильность значения $h(r_A)$.

«Бесключевой» протокол Шамира. Рассмотрим протокол, позволяющий передать ключ без использования какой-либо общей секретной информации. Этот протокол иногда называют трехпроходным протоколом Шамира-Ривеста-Адлемана.

Пусть имеется некоторое коммутирующее шифрующее преобразование E : для любого сообщения x и произвольных ключах k_1 и k_2 выполнено равенство $E_{k_1}(E_{k_2}(x)) = E_{k_2}(E_{k_1}(x))$. Тогда абоненты A и B могут реализовать следующий трехпроходной протокол для передачи секретного ключа k от A к B :

$$\begin{aligned}
&1)\ \ A \rightarrow B:\ \ E_{k_A}(k);\\
&2)\ \ A \leftarrow B:\ \ E_{k_B}(E_{k_A}(k));\\
&3)\ \ A \rightarrow B:\ \ D_{k_A}\big(E_{k_B}(E_{k_A}(k))\big) = E_{k_B}(k),
\end{aligned}$$

при этом в данном протоколе можно использовать не каждое коммутирующее преобразование E. Например, для этого протокола не подойдет преобразование $E_k(x) = k \oplus x$, так как оно заведомо нестойко в данном случае. В протоколе Шамира рекомендуется преобразование вида $E_k(x) = x^k \pmod p$, где p — большое простое число.

Данный протокол обеспечивает защиту только от пассивного противника и не обеспечивает аутентификацию.

16.1.2. Трехсторонние протоколы

Рассмотрим протоколы передачи ключей между парами участников с использованием доверенной третьей стороны T, называемой центром. В этом качестве обычно выступает неко-

торый узел сети или сервер, которому доверяют все участники. Центр T хранит ключи всех абонентов сети.

Протокол широкоротой лягушки (**Wide-Mouth-Frog**). Это простейший протокол передачи ключа k, сгенерированного участником A :

$$\begin{array}{lll} 1) & A \to T: & A,\ E_{k_{AT}}(t_A, B, k); \\ 2) & T \to B: & E_{k_{BT}}(t_T, A, k). \end{array}$$

Участник B, получив сообщение от центра T, проверяет, чтобы метка времени t_T превосходила все предыдущие метки времени, указанные центром T.

Для реализации данного протокола необходима синхронизация часов, что создает дополнительные трудности. В нем предполагается, что выбирает сеансовый ключ k и пересылает его пользователю B. Это означает, что пользователь B верит в компетентность A, в его способность создать стойкий ключ и хранить его в секрете. Такое сильное требование служит основной причиной слабого применения на практике протокола широкоротой лягушки.

Корректная временная метка означает, что сеансовый ключ был создан недавно. Однако пользователь A мог сгенерировать этот ключ годы назад и хранить его на своем жестком диске, куда противник имел возможность забраться несколько раз и снять копию ключа.

Протокол Kerberos. В основу данного протокола легли идеи протокола Нидхейма-Шредера. Протокол Kerberos имеет в настоящее время широкое распространение. Протокол Kerberos был разработан в Массачусетском технологическом институте в середине 1980-х годов и сейчас является фактическим стандартом системы централизованной аутентификации и распределения ключей симметричного шифрования. Поддерживается операционными системами семейства Unix, Windows (начиная с Windows'2000), есть реализации для Mac OS.

Централизованное распределение ключей симметричного шифрования подразумевает, что у каждого абонента сети есть только один основной ключ, который используется для взаи-

модействия с центром распределения ключей. Чтобы получить ключ шифрования для защиты обмена данными с другим абонентом, пользователь обращается к серверу ключей, который назначает этому пользователю и соответствующему абоненту сеансовый симметричный ключ.

Рассмотрим сначала базовый протокол, применяемый в протоколе аутентификации и распределения ключей Kerberos. Он состоит из следующих шагов:

$$
\begin{aligned}
&1)\ A \to T:\ A,\ B,\ r_A;\\
&2)\ A \leftarrow T:\ E_{k_{AT}}(k, r_A, L, B),\ E_{k_{BT}}(k, A, L);\\
&3)\ A \to B:\ E_{k_{BT}}(k, A, L),\ E_k(A, t, k_A);\\
&4)\ A \leftarrow B:\ E_k(t, k_B),
\end{aligned}
$$

где t — метка времени; L — период действия билета $E_{k_{BT}}(k, A, L)$; r_A — случайное число, сгенерированное участником A и вставленное в передаваемое сообщение для взаимной аутентификации; k_A, k_B — случайные числа, сгенерированные абонентами A и B соответственно и используемые либо в качестве ключа шифрования информации другой стороне, либо для выработки общего ключа $k_{AB} = f(k_A, k_B)$ с помощью некоторой функции f.

Опишем подробнее данный базовый протокол:

1. В первом обращении абонент A сообщает T, что он хотел бы связаться с B.

2. Если T разрешает эту связь, то создает билет $E_{k_{BT}}(k, A, L)$, зашифрованный ключом k_{BT}, и отсылает его A для передачи B. Абонент A получает копию этого ключа в той форме, которую он может прочесть.

3. Абонент A посылает зашифрованную временную метку t и билет $E_{k_{BT}}(k, A, L)$ участнику B.

4. Абонент B отсылает назад зашифрованную величину t, проверив, что временная метка является свежей, показывая тем самым, что он знает сеансовый ключ и готов к связи.

В полном протоколе Kerberos описанный выше базовый протокол используется два раза, так как в нем предусмотрены два сервера (рис. 16.1).

Протокол Kerberos обеспечивает распределение ключей симметричного шифрования и проверку подлинности пользователей, работающих в незащищенной сети. Предполагается, что компьютерная сеть состоит из клиентов и сервера (программная система, построенная по архитектуре «клиент-сервер»), причем клиентами могут быть пользователи, программы или специальные службы. Клиентская часть устанавливается на все компьютеры защищаемой сети, кроме тех, на которые устанавливаются компоненты сервера Kerberos.

Серверная часть Kerberos называется центром распределения ключей (Key Distribution Center, сокр. KDC) и состоит из двух компонент:

- сервер аутентификации (Authentication Server, сокр. AS);
- сервер выдачи билетов (Ticket Granting Server, сокр. TGS).

Серверная часть хранит центральную базу данных, включающую как клиентов, так и их секретные ключи. Цель данного протокола состоит в идентификации клиентов и генерировании для них сеансовых ключей. Схема функционирования протокола Kerberos представлена на рис. 16.1.

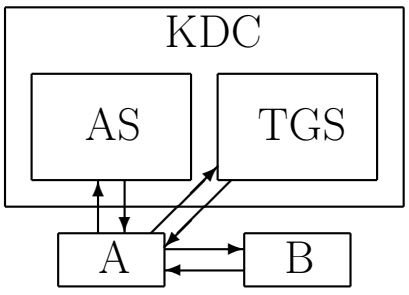


Рис. 16.1. Обмен в протоколе Kerberos

Пусть абонент A собирается начать взаимодействие с абонентом (сервером) B. Протокол предполагает следующие шаги:

1) $A \to AS$: $A,\ TGS,\ r_A$;
2) $A \leftarrow AS$: $E_{k_{A\_AS}}(k_{A\_TGS}, r_A, L_1, TGS),\ E_{k_{AS\_TGS}}(k_{A\_TGS}, A, L_1)$;
3) $A \to TGS$: $A,\ B,\ \widetilde{r}_A,\ E_{k_{AS\_TGS}}(k_{A\_TGS}, A, L_1),\ E_{k_{A\_TGS}}(A, t_1, \widetilde{r}_A)$;
4) $A \leftarrow TGS$: $E_{k_{A\_TGS}}(k, \widetilde{r}_A, L_2, B),\ E_{k_{B\_TGS}}(k, A, L_2)$;
5) $A \to B$: $E_{k_{B\_TGS}}(k, A, L_2),\ E_k(A, t_2, k_A)$;
6) $A \leftarrow B$: $E_k(t_2, k_B)$.

Опишем подробнее данный протокол:

1. Абонент A посылает серверу аутентификации AS свой идентификатор и идентификатор сервера выдачи билетов (идентификаторы передаются открытым текстом).

2. Сервер аутентификации AS, проверив, что клиент A имеется в его базе, возвращает ему билет $E_{k_{AS\_TGS}}(k_{A\_TGS}, A, L_1)$ для доступа к серверу выдачи билетов, ключ k_{A_TGS} для взаимодействия с сервером выдачи билетов TGS и период L_1 действия билета.

3. Абонент A на этот раз обращается к серверу выдачи билетов TGS. Он пересылает полученный от AS билет:

$$E_{k_{AS\_TGS}}(k_{A\_TGS}, A, L_1)$$

и аутентификационный блок $E_{k_{A\_TGS}}(A, t_1, \widetilde{r}_A)$, содержащий идентификатор A и метку времени t_1, показывающую, когда была сформировано сообщение. Сервер выдачи билетов расшифровывает полученный билет $E_{k_{AS\_TGS}}(k_{A\_TGS}, A, L_1)$ и получает из него информацию о том, кому был выдан билет, на какой срок и ключ шифрования k_{A_TGS}, сгенерированный сервером AS для взаимодействия между абонентом A и сервером TGS. С помощью этого ключа расшифровывается аутентификационный блок. Если идентификатор в блоке совпадает с идентификатором в билете, это доказывает, что посылку сгенерировал на самом деле A. Далее делается проверка времени действия билета L_1 и времени t_1 отправления сообщения. Если проверка проходит и действующая в системе

политика позволяет абоненту A обращаться к абоненту B, тогда выполняется следующий шаг.

4. Сервер выдачи билетов TGS посылает абоненту A ключ шифрования k и билет $E_{k_{B\_TGS}}(k, A, L_2)$, необходимые для обращению к абоненту (серверу) B.

5. Абонент A посылает билет $E_{k_{B\_TGS}}(k, A, L_2)$, полученный от сервера выдачи билетов, и свой аутентификационный блок абоненту B, с которым хочет установить сеанс защищенного взаимодействия. Предполагается, что B уже зарегистрировался в системе и распределил с сервером TGS ключ шифрования k_{B_TGS}. Имея этот ключ, он может расшифровать билет, получить ключ шифрования k и проверить подлинность отправителя сообщения.

6. Теперь уже B должен доказать A свою подлинность. Он может сделать это, показав, что правильно расшифровал предыдущее сообщение. Вот поэтому, B берет отметку времени t_2 из аутентификационного блока A, шифрует на ключе k и возвращает A.

Благодаря введению второго сервера нагрузка на первый сервер уменьшается во много раз. Сервер аутентификации AS должен быть наиболее защищенным, поскольку он хранит главные ключи всех пользователей. Серверов выдачи билетов TGS может быть несколько.

Семейство протоколов KriptoKnight

Семейство протоколов KriptoKnight было задумано как гибкий и компактный набор протоколов для различных сценариев аутентифицированного обмена ключами.

Протокол взаимной аутентификации 2PAP (Two Party Key Distribution Protocol):

$$
\begin{aligned}
&1)\ \ A \to B:\ r_A;\\
&2)\ \ A \leftarrow B:\ r_B,\ h_{k_{AB}}(r_A, r_B, B);\\
&3)\ \ A \to B:\ h_{k_{AB}}(r_A, r_B).
\end{aligned}
$$

Двусторонний аутентифицированный протокол получения ключа 2PAKDP (Two Party Authenticated Key Distribution Protocol):

1) $A \to B:\ r_A$;
2) $A \leftarrow B:\ h_{k_{AB}}(r_A, k, B),\ k \oplus E_{k_{AB}}(h_{k_{AB}}(r_A, k, B))$;
3) $A \to B:\ [h_{k_{AB}}(r_A, k)]$,

где квадратные скобки означают необязательность сообщения. На втором шаге протокола участник A сначала зашифровывает значение $h_{k_{AB}}(r_A, k, B)$ с помощью симметричного шифра $E_{k_{AB}}$ на ключе k_{AB} и вычисляет ключ k. Затем он получает свертку сообщения (r_A, k, B) с помощью хеш-функции h и проверяет, что сообщение получено именно от B и правильность получения ключа k.

Трехсторонний аутентифицированный протокол получения ключа 3PAKDP (Tree Party Authenticated Key Distribution Protocol):

1) $A \to B:\ r_A$;
2) $B \to T:\ r_A,\ r_B,\ A$;
3) $A \leftarrow T:\ h_{k_{AT}}(r_A, k, B),\ k \oplus E_{k_{AT}}(h_{k_{AT}}(r_A, k, B))$;
4) $B \leftarrow T:\ h_{k_{BT}}(r_B, k, A),\ k \oplus E_{k_{BT}}(h_{k_{BT}}(r_B, k, A))$.

В данном случае абоненты A и B на третьем и четвертом шаге протокола соответственно проводят те же операции, что абонент A на втором шаге предыдущего протокола.

16.2. Передача ключей с использованием асимметричного шифрования

Протоколы без использования электронной подписи

Для передачи ключа k можно использовать следующий одношаговый протокол:

$$A \to B:\ E_B(k, t, A),$$

где E_B — алгоритм шифрования с открытым ключом абонента B, t — метка времени.

Протокол NSPK. Данный протокол имеет вид:

$$1)\ A \to B:\ E_B(k_A, A);$$
$$2)\ A \leftarrow B:\ E_A(k_A, k_B);$$
$$3)\ A \to B:\ E_B(k_B),$$

где k_A и k_B — ключевые переменные. В данном протоколе каждая из сторон генерирует свое значение ключа, которое можно использовать для вычисления общего ключа от значений k_A и k_B. После расшифрования полученных сообщений на втором и третьем шагах и после проверки совпадения значений k_A и k_B стороны убеждаются в том, что они имеют дело именно с нужной стороной и что другая сторона правильно расшифровала полученное значение ключа.

Протоколы с использованием электронной подписи

При использовании электронной подписи аутентифицированный протокол передачи ключей может содержать только одно сообщение и иметь один из следующих трех видов:

1) зашифрование и подпись ключа:

$$A \to B:\ E_B(k, t),\ Sig_A(B, k, t);$$

2) зашифрование подписанного ключа:

$$A \to B:\ E_B(k, t, Sig_A(B, k, t));$$

3) подпись зашифрованного ключа:

$$A \to B:\ t,\ E_B(A, k),\ Sig_A(B, t, E_B(A, k)),$$

где $Sig_A(m)$ — результат применения алгоритма формирования электронной подписи к сообщению m при секретном ключе участника A.

Сертификаты открытых ключей

Центры сертификации осуществляют проверку и подтверждение подлинности открытых ключей. Как правило, при использовании открытых ключей хранятся и пересылаются не сами ключи, а их сертификаты. Центры сертификации выдают

сертификаты, осуществляют проверку их подлинности и ведут списки отозванных сертификатов.

Сертификат представляет собой набор данных:

$$cert_A = (A, k_A, t, Sig_T(A, k_A, t)),$$

состоящий из идентификатора абонента A, его открытого ключа k_A и дополнительной информации, включающей время t выдачи сертификата, срок его действия, предназначение ключа, заверенный электронной подписью доверенного центра T или заслуживающего доверия лица. Сертификат предназначен для исключения возможности подмены открытого ключа при его хранении или пересылке.

Получив такой сертификат и проверив электронную подпись, можно убедиться в том, что открытый ключ действительно принадлежит данному абоненту.

Международный стандарт МККТТ X.509 (ISO 9594-8) определяет следующий протокол идентификации с одновременным распределением ключей:

$$\begin{array}{ll} 1) & A \to B: \ cert_A, \ d_A, \ Sig_A(d_A); \\ 2) & A \leftarrow B: \ cert_B, \ d_B, \ Sig_B(d_B); \\ 3) & A \to B: \ r_B, \ B, \ Sig_A(r_B, B), \end{array}$$

где $cert_A$, $cert_B$ — сертификаты сторон; Sig_A, Sig_B — алгоритмы формирования электронных подписей сторон; d_A, d_B — наборы передаваемых и заверяемых электронной подписью данных:

$$d_A = (t_A, r_A, B, data_1, E_B(k_A)),$$

$$d_B = (t_B, r_B, A, r_A, data_2, E_A(k_B)).$$

В поле $data$ заносится дополнительная информация для аутентификации источника.

16.3. Открытое распределение ключей

Открытое распределение ключей представляет собой протокол обмена сообщениями по открытому каналу связи, позволяющий участникам после завершения обмена выработать общий

секретный ключ. Важным преимуществом открытого распределения ключей является то, что ни один из абонентов заранее не может предугадать значение ключа, так как ключ зависит от содержания сообщений, передаваемых в процессе обмена. Но более важным является то обстоятельство, что участники вырабатывают ключ без какой-либо общей секретной информации, распределяемой заранее.

Основным протоколом открытого распределения ключей является протокол Диффи-Хеллмана, а все различие заключается в способах его усиления для повышения защищенности к известным атакам.

Протокол Диффи-Хеллмана и его усиления

Протокол DH. Первый алгоритм открытого распределения ключей был предложен в 1976 г. У. Диффи и М. Хеллманом. Пусть p — достаточно большое простое число, g — первообразный корень по модулю p (образующий элемент мультипликативной группы $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$). Абоненты A и B генерируют случайные числа x_A, $1 \leqslant x_A \leqslant p-2$, и x_B, $1 \leqslant x_B \leqslant p-2$, соответственно. Данные числа каждым абонентом хранятся в секрете. Далее они должны обменяться сообщениями в соответствии с протоколом:

$$1)\ \ A \to B:\ g^{x_A} \pmod p;$$
$$2)\ \ A \leftarrow B:\ g^{x_B} \pmod p.$$

После этого искомый ключ вычисляется по формуле:

$$k = (g^{x_A})^{x_B} = (g^{x_B})^{x_A} \pmod p.$$

Недостатком этого протокола является возможность атаки типа «противник в середине», при которой противник E может осуществить подмену передаваемых абонентами сообщений. В этом случае противник E фиксирует некоторые числа $x_{AE}, x_{BE} \in \{1, \ldots, p-2\}$ и подменяет сообщения следующим

образом:

$$\begin{aligned} A \to E(B):&\ \ g^{x_A} \pmod p;\\ E \to B:&\ \ g^{x_{BE}} \pmod p;\\ E(A) \leftarrow B:&\ \ g^{x_B} \pmod p;\\ A \leftarrow E:&\ \ g^{x_{AE}} \pmod p. \end{aligned}$$

После этого E вычисляет значения ключей:

$$k_{AE} = (g^{x_A})^{x_{AE}} \pmod p, \quad k_{BE} = (g^{x_B})^{x_{BE}} \pmod p$$

для связи с пользователями A и соответственно B. В результате этого противник получает возможность полностью контролировать обмен сообщениями между абонентами A и B.

Предварительное распределение. Предположим, что имеется доверенный центр, который может с использованием сертификатов связать идентификаторы участников со значениями вида $g^x \pmod p$, которыми участники обменивались в протоколе DH. Каждый участник должен предварительно получить в центре сертификат вида

$$cert_A = (A, y_A, Sig_T(A, y_A)),$$

где $y_A = g^{x_A} \pmod p$, Sig_T — подпись доверенного центра T.

Теперь для выработки совместного значения ключа участники обмениваются сертификатами:

$$\begin{aligned} 1)&\ \ A \to B:\ \ cert_A;\\ 2)&\ \ A \leftarrow B:\ \ cert_B. \end{aligned}$$

Такой протокол защищен от атаки «противник в середине», но вместе с тем формируемый ключ остается неизменным, поэтому исчезают все преимущества открытого распределения.

Семейство протоколов MTI. Протоколы MTI были предложены Т. Мацумото, И. Такашима и Х. Имаи для защиты протокола DH от атаки «противник в середине».

Пусть пользователи A и B имеют соответствующие секретные ключи x_A $(1 \leqslant x_A \leqslant p-2)$ и x_B $(1 \leqslant x_B \leqslant p-2)$ и публикуют свои открытые ключи $y_A = g^{x_A} \pmod p$, $y_B = g^{x_B} \pmod p$. В некоторых из представленных ниже протоколах используются

x_A^{-1}, x_B^{-1}. В этом случае должны выполняться условия:

$$(x_A, p-1) = (x_B, p-1) = 1,$$

$$x_A x_A^{-1} \equiv 1 \pmod{p-1}, \quad x_B x_B^{-1} \equiv 1 \pmod{p-1}.$$

Также пользователи A и B должны сгенерировать соответствующие случайные числа α $(1 \leqslant \alpha \leqslant p-2)$ и β $(1 \leqslant \beta \leqslant p-2)$.

MTI/A0. Для выработки общего секретного ключа k пользователи A и B обмениваются следующими сообщениями:

$$A \to B: \ g^{\alpha} \pmod{p};$$
$$A \leftarrow B: \ g^{\beta} \pmod{p}.$$

Затем участники вычисляют следующие значения:

$$A: \ k_A = (g^{\beta})^{x_A} y_B^{\alpha} = g^{\beta x_A + \alpha x_B},$$

$$B: \ k_B = (g^{\alpha})^{x_B} y_A^{\beta} = g^{\alpha x_B + \beta x_A}.$$

Значение $k = k_A = k_B$ и будет общим ключом.

MTI/A(s). Для выработки общего секретного ключа k пользователи A и B обмениваются следующими сообщениями:

$$A \to B: \ g^{\alpha x_A^s} \pmod{p};$$
$$A \leftarrow B: \ g^{\beta x_B^s} \pmod{p}.$$

Затем участники вычисляют следующие значения:

$$A: \ k_A = (g^{\beta x_B^s})^{x_A} y_B^{\alpha x_A^s} = g^{\beta x_A x_B^s + \alpha x_B x_A^s},$$

$$B: \ k_B = (g^{\alpha x_A^s})^{x_B} y_A^{\beta x_B^s} = g^{\alpha x_A^s x_B + \beta x_B^s x_A}.$$

Значение $k = k_A = k_B$ и будет общим ключом.

MTI/B0. Участники обмениваются сообщениями:

$$A \to B: \ y_B^{\alpha} \pmod{p};$$
$$A \leftarrow B: \ y_A^{\beta} \pmod{p}$$

и вычисляют значения

$$A: \ k_A = (y_A^{\beta})^{x_A^{-1}} g^{\alpha} = g^{\alpha+\beta},$$

$$B: \ k_B = (y_B^{\alpha})^{x_B^{-1}} g^{\beta} = g^{\alpha+\beta}.$$

MTI/B(s). Участники обмениваются сообщениями:

$$A \to B:\ y_B^{\alpha x_A^s} \pmod p;$$
$$A \leftarrow B:\ y_A^{\beta x_B^s} \pmod p$$

и вычисляют значения:

$$A:\ k_A = (y_A^{\beta x_B^s})^{x_A^{-1}} g^{\alpha x_A^s} = g^{\alpha x_A^s + \beta x_B^s},$$

$$B:\ k_B = (y_B^{\alpha x_A^s})^{x_B^{-1}} g^{\beta x_B^s} = g^{\alpha x_A^s + \beta x_B^s}.$$

MTI/C0. Участники обмениваются сообщениями:

$$A \to B:\ y_B^{\alpha} \pmod p;$$
$$A \leftarrow B:\ y_A^{\beta} \pmod p$$

и вычисляют значения:

$$A:\ k_A = (y_A^{\beta})^{x_A^{-1}\alpha} = g^{\alpha\beta},$$

$$B:\ k_B = (y_B^{\alpha})^{x_B^{-1}\beta} = g^{\alpha\beta}.$$

MTI/C(s). Участники обмениваются сообщениями:

$$A \to B:\ y_B^{\alpha x_A^s} \pmod p;$$
$$A \leftarrow B:\ y_A^{\beta x_B^s} \pmod p$$

и вычисляют значения:

$$A:\ k_A = (y_A^{\beta x_B^s})^{x_A^{-1}\alpha x_A^s} = g^{\alpha\beta x_A^s x_B^s},$$

$$B:\ k_B = (y_B^{\alpha x_A^s})^{x_B^{-1}\beta x_B^s} = g^{\alpha\beta x_A^s x_B^s}.$$

Модифицированное семейство протоколов MTI на эллиптических кривых

Любая криптосистема, основанная на дискретном логарифмировании, легко может быть перенесена на эллиптические кривые. В этом случае операция $y = g^x \pmod p$ заменяется на $Y = [x]G \pmod p$. Пусть q — некоторый (достаточно большой) простой делитель числа $|E_p(a,b)|$, где $E_p(a,b)$ — эллиптическая кривая над полем $\mathbf{Z}_p$ вида:

$$E_p(a,b):\ y^2 = x^3 + ax + b \pmod p.$$

Пусть некоторая точка $G \in E_p(a,b)$ имеет порядок q, т.е. образует циклическую подгруппу порядка q в $(E_p(a,b),+)$:

$$\langle G\rangle = \{G,\ [2]G, \ldots, [q]G = \mathcal{O}\}.$$

Общедоступные параметры системы: p, q, G, $E_p(a,b)$. Абоненты A и B выбирают соответствующие секретные ключи x_A и x_B, не превосходящие числа $q-1$. По каждому секретному ключу вычисляется открытый ключ:

$$Y_A = [x_A]G,$$
$$Y_B = [x_B]G.$$

В некоторых из представленных ниже протоколах используются x_A^{-1}, x_B^{-1}. В этом случае должно выполняться условие:

$$(x_A, q) = (x_B, q) = 1,$$
$$x_A x_A^{-1} \equiv 1 \pmod q, \quad x_B x_B^{-1} \equiv 1 \pmod q.$$

Также пользователи A и B должны сгенерировать соответствующие случайные числа α $(1 \leqslant \alpha \leqslant q-1)$ и β $(1 \leqslant \beta \leqslant q-1)$.

Модификация протокола MTI/A0. Для выработки общего секретного ключа k пользователи A и B обмениваются следующими сообщениями:

$$A \to B:\ [\alpha]G;$$
$$A \leftarrow B:\ [\beta]G.$$

Затем участники вычисляют следующие точки эллиптической кривой $E_p(a,b)$:

$$Z_A = [x_A]([\beta]G) + [\alpha]Y_B = [\beta x_A + \alpha x_B]G,$$
$$Z_B = [x_B]([\alpha]G) + [\beta]Y_A = [\alpha x_B + \beta x_A]G.$$

При этом $Z_A = Z_B$. Теперь абоненты A и B могут использовать, например абсциссу точки Z_A, в качестве ключа k для секретной переписки.

Модификация протокола MTI/A(s). Для выработки общего секретного ключа k пользователи A и B обмениваются следующими сообщениями:

$$A \to B:\ [\alpha x_A^s]G;$$
$$A \leftarrow B:\ [\beta x_B^s]G.$$

Затем участники вычисляют следующие значения:

$$Z_A = [x_A]([\beta x_B^s]G) + [\alpha x_A^s]Y_B = [\beta x_A x_B^s + \alpha x_B x_A^s]G,$$

$$Z_B = [x_B]([\alpha x_A^s]G) + [\beta x_B^s]Y_A = [\alpha x_A^s x_B + \beta x_B^s x_A]G.$$

Модификация протокола MTI/B0. Участники обмениваются сообщениями:

$$A \to B: \ [\alpha]Y_B;$$
$$A \leftarrow B: \ [\beta]Y_A$$

и вычисляют следующие точки эллиптической кривой $E_p(a, b)$:

$$Z_A = [x_A^{-1}]([\beta]Y_A) + [\alpha]G = [\alpha + \beta]G,$$

$$Z_B = [x_B^{-1}]([\alpha]Y_B) + [\beta]G = [\alpha + \beta]G.$$

Модификация протокола MTI/B(s). Участники обмениваются сообщениями:

$$A \to B: \ [\alpha x_A^s]Y_B;$$
$$A \leftarrow B: \ [\beta x_B^s]Y_A$$

и вычисляют значения:

$$Z_A = [x_A^{-1}]([\beta x_B^s]Y_A) + [\alpha x_A^s]G = [\alpha x_A^s + \beta x_B^s]G,$$

$$Z_B = [x_B^{-1}]([\alpha x_A^s]Y_B) + [\beta x_B^s]G = [\alpha x_A^s + \beta x_B^s]G.$$

Модификация протокола MTI/C0. Участники обмениваются сообщениями:

$$A \to B: \ [\alpha]Y_B;$$
$$A \leftarrow B: \ [\beta]Y_A$$

и вычисляют следующие точки эллиптической кривой $E_p(a, b)$:

$$Z_A = [x_A^{-1}\alpha]([\beta]Y_A) = [\alpha\beta]G,$$

$$Z_B = [x_B^{-1}\beta]([\alpha]Y_B) = [\alpha\beta]G.$$

Модификация протокола MTI/C(s). Участники обмениваются сообщениями:

$$A \to B: \ [\alpha x_A^s]Y_B;$$
$$A \leftarrow B: \ [\beta x_B^s]Y_A$$

и вычисляют следующие точки эллиптической кривой $E_p(a,b)$:

$$Z_A = [x_A^{-1}\alpha x_A^s]([\beta x_B^s]Y_A) = [\alpha\beta x_A^s x_B^s]G,$$
$$Z_B = [x_B^{-1}\beta x_B^s]([\alpha x_A^s]Y_B) = [\alpha\beta x_A^s x_B^s]G.$$

Открытое распределение ключей с использованием самосертифицируемых ключей. Схема, предложенная М. Гиролтом, основана на схеме RSA с простыми числами вида $p = 2p_1 + 1$, $q = 2q_1 + 1$, где p_1, q_1 — простые числа. Пусть g — элемент порядка $2p_1q_1 = (p-1, q-1)$ мультипликативной группы $\mathbb{Z}_n^*$, $n = pq$. Пусть также $de = 1 \pmod n$, причем d — секретная экспонента, известная только автору выдачи сертификата, e — открытая (d и e общие для всех абонентов).

Абонент A генерирует свое секретное число x_A, для которого выполнено $1 < x_A < 2p_1q_1$, вычисляет открытое значение $y_A = g^{\alpha_A}$, представляет в центр y_A и получает в центре значение $p_A = (y_A - A)^d \pmod n$ (идентификатор абонента и его имя обозначены одним символом). По значению p_A можно убедиться, что оно принадлежит абоненту A : значение $(p_A^e + A) \pmod n$ должно совпасть с y_A. Поэтому p_A можно использовать в качестве сертификата.

Протокол выработки ключа имеет такой вид:

$$1)\ \ A \to B:\ \ A,\ p_A,\ \gamma_A = g^{r_A} \pmod n;$$
$$2)\ \ A \leftarrow B:\ \ B,\ p_B,\ \gamma_B = g^{r_B} \pmod n.$$

Общий ключ $k_{AB} = g^{r_A x_B + r_B x_A} \pmod n$ вычисляется теперь участниками A и B по формулам:

$$k_{AB} = \gamma_A^{x_B}(p_A^e + A)^{r_B} = \gamma_B^{x_A}(p_B^e + B)^{r_A} \pmod n.$$

Протокол «унифицированная модель». Данный протокол предложен Р. Анни, Д. Джонсоном и М. Матиасом в 1995 г. В нем используются сертификаты открытых ключей. Пусть α и β — случайные числа, сгенерированные абонентами A и B соответственно (как и в протоколах MTI). Данные абоненты обмениваются следующими сообщениями:

$$A \to B:\ \ cert_A,\ g^{\alpha} \pmod n;$$
$$A \leftarrow B:\ \ cert_B,\ g^{\beta} \pmod n.$$

Общий ключ $k = h(g^{x_A x_B}, g^{\alpha\beta})$ вычисляется участниками A и B по формулам:

$$k = h(y_B^{x_A}, g^{\beta\alpha}) = h(y_A^{x_B}, g^{\alpha\beta}),$$

где h — бесключевая хеш-функция.

16.4. Предварительное распределение ключей

Для предварительного распределения ключей стороны могут обменяться ключами при личной встрече либо использовать некоторый защищенный канал. Иногда же оказывается удобным распределять не сами ключи, а некоторые вспомогательные ключевые материалы, на основании которых каждый участник может самостоятельно вычислить ключ, используя для этого некоторую установленную заранее процедуру.

Схема Блома. Пусть F — конечное поле. Для n абонентов некоторой сети зафиксируем n попарно различных значений $r_1, \ldots, r_n \in F \setminus \{0\}$. Элемент r_i припишем абоненту A_i, $i = 1, \ldots, n$. Это элементы не являются секретными и могут храниться на общедоступном сервере сети. Выберем симметрический многочлен $f(x, y) \in F[x, y]$ степени m $(1 \leqslant m \leqslant n-1)$:

$$f(x, y) = \sum_{i=0}^{m} \sum_{j=0}^{m} a_{ij} x^i y^j, \quad a_{ij} = a_{ji}, \quad 0 \leqslant i \leqslant j \leqslant m.$$

Коэффициенты a_{ij} многочлена $f(x, y)$ являются секретными и должны храниться только в центре распределения ключей. Каждый абонент получает в качестве ключевых материалов набор $q_i = (b_{i0}, \ldots, b_{im})$, состоящий из коэффициентов многочлена:

$$g_i(x) = f(r_i, x) = b_{i0} + b_{i1} x + \ldots + b_{im} x^m.$$

Для связи между абонентами A_i и A_j теперь можно использовать общий ключ k_{ij} :

$$k_{ij} = g_i(r_j) = f(r_i, r_j) = f(r_j, r_i) = g_j(r_i) = k_{ji}.$$

Заметим, что при использовании данной схемы каждый абонент хранит $m + 1$ секретных значений вместо $n - 1$.

Говорят, что схема предварительного распределения ключей является стойкой к m-кратной компрометации ключей (к сговору m абонентов), если после того как противнику станут известны ключевые материалы m абонентов, он не сможет получить никакой информации о ключах парной связи остальных абонентов.

Теорема 16.1. Схема Блома предварительного распределения ключей между n абонентами, использующая многочлен степени m, $1 \leqslant m \leqslant n-1$, является стойкой к m-кратной компрометации ключей.

Доказательство. Заметим, что матрица всех ключей вычисляется по следующей формуле:

$$\begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{pmatrix} =$$

$$\begin{pmatrix} 1 & r_1 & r_1^2 & \dots & r_1^m \\ 1 & r_2 & r_2^2 & \dots & r_2^m \\ \dots & \dots & \dots & \dots & \dots \\ 1 & r_n & r_n^2 & \dots & r_n^m \end{pmatrix} \begin{pmatrix} a_{00} & a_{01} & \dots & a_{0m} \\ a_{10} & a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots & \dots \\ a_{m0} & a_{m1} & \dots & a_{mm} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_n \\ r_1^2 & r_2^2 & \dots & r_n^2 \\ \dots & \dots & \dots & \dots \\ r_1^m & r_2^m & \dots & r_n^m \end{pmatrix}.$$

Поэтому ключ k_{ij} вычисляется по следующей формуле:

$$k_{ij} = \begin{pmatrix} 1 & r_i & r_i^2 & \dots & r_i^m \end{pmatrix} \begin{pmatrix} a_{00} & a_{01} & \dots & a_{0m} \\ a_{10} & a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots & \dots \\ a_{m0} & a_{m1} & \dots & a_{mm} \end{pmatrix} \begin{pmatrix} 1 \\ r_j \\ r_j^2 \\ \dots \\ r_j^m \end{pmatrix}.$$

Здесь матрица $A = (a_{ij})$ размера $(m+1) \times (m+1)$ составлена из коэффициентом многочлена $f(x, y)$ и является симметрической.

Предположим, что известны ключевые материалы, принадлежащие m абонентам. Покажем, что по ним нельзя получить никакой информации ни об одном из ключей для связи между оставшимися $n-m$ абонентами. Без ограничения общности

можно считать, что известны секретные наборы ключевых материалов первых m абонентов: $A_1, \ldots, A_m$. Используя эту информацию, будем искать ключ для связи между абонентами A_i и A_j, $m+1 \leqslant i < j \leqslant n$. Рассмотрим равенство:

$$\left(\begin{array}{cccc|c} k_{11} & k_{12} & \ldots & k_{1m} & k_{1j} \\ k_{21} & k_{22} & \ldots & k_{2m} & k_{2j} \\ \ldots & \ldots & \ldots & \ldots & \\ k_{m1} & k_{m2} & \ldots & k_{mm} & k_{mj} \\ \hline k_{i1} & k_{i2} & \ldots & k_{im} & k_{ij} \end{array}\right) =$$

$$\left(\begin{array}{ccccc} 1 & r_1 & r_1^2 & \ldots & r_1^m \\ 1 & r_2 & r_2^2 & \ldots & r_2^m \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & r_m & r_m^2 & \ldots & r_m^m \\ \hline 1 & r_i & r_i^2 & \ldots & r_i^m \end{array}\right) A \left(\begin{array}{cccc|c} 1 & 1 & \ldots & 1 & 1 \\ r_1 & r_2 & \ldots & r_m & r_j \\ r_1^2 & r_2^2 & \ldots & r_m^2 & r_j^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_1^m & r_2^m & \ldots & r_n^m & r_j^m \end{array}\right),$$

в котором в левой части неизвестен только искомый ключ k_{ij}, а в правой части неизвестны все элементы матрицы A. Запишем это равенство в виде $K = SAT$. Заметим, что в правой части последнего равенства на матрицу A слева и справа умножаются обратимые матрицы, так как все значения r_1, $\ldots$, r_n являются элементами поля F и попарно различны. Поэтому $A = S^{-1}KT^{-1}$. Нетрудно видеть, что для любых s и t, $0 \leqslant s \leqslant m$, $0 \leqslant t \leqslant m$, элемент a_{st} зависит от неизвестного ключа k_{ij}. В силу того, что **все** элементы матрицы A зависят от неизвестного ключа k_{ij}, ключевые материалы абонентов A_1, $\ldots$, A_m не дают никакой информации ни о каком элементе матрицы A, следовательно, о ключе k_{ij}. $\square$

Литература

[1] Арнольд В.И. Цепные дроби. М.: МЦНМО, 2009. 40 с.

[2] Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В. Основы криптографии. М.: Гелиос АРВ, 2005. 480 с.

[3] Бабаш А.В., Глухов М.М., Шанкин Г.П. О преобразованиях множества слов в конечном алфавите, не размножающих искажений. // Дискретная математика. 1997. Т. 9. № 3. С. 3–19.

[4] Бабаш А.В., Шанкин Г.П. Криптография. М.: СОЛОН-ПРЕСС, 2007. 512 с.

[5] Болотов А.А., Гашков С.Б., Фролов А.Б., Часовских А.А. Элементарное введение в эллиптическую криптографию: Алгебраические и алгоритмические основы. М.: КомКнига, 2006. 328 с.

[6] Болотов А.А., Гашков С.Б., Фролов А.Б. Элементарное введение в эллиптическую криптографию: Протоколы криптографии на эллиптических кривых. М.: КомКнига, 2006. 280 с.

[7] Бухштаб А.А. Теория чисел. СПб.: Лань, 2008. 383 с.

[8] Василенко О.Н. Теоретико-числовые алгоритмы в криптографии. М.: изд-во МЦНМО, 2003. 328 с.

[9] Введение в криптографию. / Под общ. ред. В.В. Ященко. 4-е изд., доп. М.: МЦНМО, 2012. 348 с.

[10] Виноградов И.М. Основы теории чисел. СПб.: Лань, 2006. 176 с.

[11] Галочкин А.И, Нестеренко Ю.В, Шидловский А.Б. Введение в теорию чисел. М.: Изд-во Моск. Ун-та, 1984. 152 с.

[12] Гашков С.Б. Упрощенное обоснование вероятностного теста Миллера–Рабина для проверки простоты чисел. // Дискрет. матем. 1998. Т. 10. № 4. С. 35–38.

[13] Глухов М.М., Елизаров В.П., Нечаев А.А. Алгебра: в 2 т. М.: Гелиос АРВ, 2003.

[14] Глухов М.М., Круглов И.А., Пичкур А.Б., Черемушкин А.В. Введение в теоретико-числовые методы криптографии. СПб.: издательство «Лань», 2011. 400 с.

[15] ГОСТ Р 34.10-2012. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи. М.: Стандартинформ, 2012.

[16] ГОСТ Р 34.11-2012. Информационная технология. Криптографическая защита информации. Функция хэширования. М.: Стандартинформ, 2013.

[17] ГОСТ Р 34.12-2015. Информационная технология. Криптографическая защита информации. Блочные шифры. М.: Стандартинформ, 2016.

[18] ГОСТ Р 34.13-2015. Информационная технология. Криптографическая защита информации. Режимы работы блочных шифров. М.: Стандартинформ, 2016.

[19] Демьянчук А.А, Мирин А.Ю., Молдовян Н.А. Типы и приложения протоколов с нулевым разглашением секрета. // Информационно-управляющие системы. 2013. № 3. С. 67–73.

[20] Ерош И.Л. Дискретная математика. Математические вопросы криптографии: учеб. пособие. СПб.: СПбГУАП, 2001. 56 с.

[21] Запечников С.В. Криптографические протоколы и их применение в финансовой и коммерческой деятельности: учеб. пособие для вузов. М.: Горячая линия – Телеком, 2007. 320 с.

[22] Зубов А.Ю. Криптографические методы защиты информации. Совершенные шифры. М.: Гелиос АРВ, 2005. 192 с.

[23] Кнэпп Э. Эллиптические кривые. М.: Факториал Пресс, 2004. 488 с.

[24] Коблиц Н. Курс теории чисел и криптографии. М.: ТВП, 2001. 254 с.

[25] Кострикин А.И. Введение в алгебру в 3-х частях. М.: ФИЗМАТЛИТ, 2001.

[26] Маховенко Е.Б. Теоретико-числовые методы в криптографии. Учеб. пособие для вузов. М.: Гелиос АРВ, 2006. 320 с.

[27] Молдовян А.А., Молдовян Д.Н., Левина А.Б. Протоколы аутентификации с нулевым разглашением секрета. СПб: университет ИТМО, 2016. 55 с.

[28] Нестеренко Ю.В. Теория чисел: Учеб. для студ. высш. учеб. заведений. М.: издательский центр «Академия», 2008. 272 с.

[29] Нестеренко А.Ю. Теоретико-числовые методы в криптографии: учеб. пособие. Моск. гос. ин-т электроники и математики. 2012. 224 с.

[30] Онацкий. А.В. Модификация протоколов Шнорра и Окамото на эллиптических кривых. // Восточно-Европейский журнал передовых технологий. 2013. Т. 66. № 6/9. С 14–18.

[31] Острик В.В., Цфасман М.А. Алгебраическая геометрия и теория чисел: рациональные и эллиптические кривые. М.: МЦНМО, 2001. 64 с.

[32] Погорелов Б.А., Сачков В.Н. Словарь криптографических терминов. М.: МЦНМО, 2006. 91 с.

[33] Прахар К. Распределение простых чисел. М.: Мир, 1967. 511 с.

[34] Прохоров Ю.Г. Эллиптические кривые и криптография. Семестр 1. М.: МГУ, 2007. 144 с.

[35] Рацеев С.М. Некоторые обобщения теории Шеннона о совершенных шифрах. // Вестн. ЮУрГУ. Сер. Матем. моделирование и программирование. 2015. Т. 8. № 1. С. 111–127.

[36] Рацеев С.М. Элементы криптографии. Часть 1. Ульяновск: УлГУ, 2012. 112 с.

[37] Рацеев С.М. Элементы криптографии. Часть 2. Ульяновск: УлГУ, 2013. 116 с.

[38] Рацеев С.М., Череватенко О.И. О кодах аутентификации на основе ортогональных таблиц. // Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки. 2014. Т. 37. № 4. С. 178–186.

[39] Рид М. Алгебраическая геометрия для всех. М.: Мир, 1991. 151 с.

[40] Ростовцев А.Г., Маховенко Е.Б. Теоретическая криптография. СПб.: АНО НПО ПРОФЕССИОНАЛ, 2005. 490 с.

[41] Рябко Б.Я., Фионов А.Н. Криптографические методы защиты информации. М.: Горячая линия – Телеком, 2005. 229 с.

[42] Саломаа А. Криптография с открытым ключом. М.: Мир, 1996. 318 с.

[43] Фомичев В.М. Методы дискретной математики в криптологии. М.: Диалог-МИФИ, 2010. 424 с.

[44] Харин Ю.С., Берник В.И., Матвеев Г.В. Математические основы криптологии. Минск: БГУ, 1999. 319 с.

[45] Холл М. Комбинаторика. М.: Мир, 1970. 424 с.

[46] Черемушкин А.В. Криптографические протоколы. Основные свойства и уязвимости. М.: Академия, 2009. 272 с.

[47] Шафаревич И.Р. Основы алгебраической геометрии. Алгебраические многообразия в проективном пространстве. Т. 1. М.: Наука, 1988. 183 с.

[48] Alford W.R., Granville A., Pomerance C. There are infinitely many Carmichael numbers. // Ann. Math. 1994. Vol. 140. P. 703–722.

[49] Bose R.S. On the applications of the properties of Galois fields to the problems of construction of Hyper-Graeco-Latin squares. // Indian J. Stat. 1938. № 3, part 4. P. 323–338.

[50] Chor B., Rivest R.L. A Knapsack-Type Public Key Cryptosystem Based on Arithmetic in Finite Fields. // Transactions on information theory. 1988. Vol. 34, № 5. P. 901–909.

[51] Announcing the ADVANCED ENCRYPTION STANDARD (AES). Federal Information Processing Standards Publication 197. United States National Institute of Standards and Technology (NIST). November 26, 2001. Retrieved October 2, 2012.

[52] An Elliptic Curve Cryptography (ECC). Primer why ECC is the next generation of public key cryptography. The Certicom 'Catch the Curve' White Paper Series, June 2004. 24 p.

[53] Blum M. Coin flipping by telephone: A protocol for solving impossible problems. // Proc. 24th IEEE Comp. Conf., 1982. P. 133–137; reprinted in SIGACT News, Vol. 15, № 1, 1983. P. 23–27.

[54] Chaum D., Pedersen T. P. Wallet databases with observers. // Proc. Crypto'92, Lect. Notes in Comput. Sci. 1993. Vol. 740. P. 89–105.

[55] Cramer R., Gennaro R., Schoenmakers B. A secure and optimally efficient multi-authority election scheme. // Proc.

EUROCRYPT'97, Lect. Notes in Comput. Sci. 1997. Vol. 1233. P. 103–118.

[56] Cramer R., Franklin M., Schoenmakers B., Yung M. Multi-Authority Secret-Ballot Elections with Linear Work. // Proc. EUROCRYPT'96, Lect. Notes in Comput. Sci. 1996. Vol. 1070. P. 72–83.

[57] Goldreich O., Micali S., Wigderson A. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. // J. ACM. 1991. Vol. 38, № 3. P. 691–729.

[58] Girault M. An identity-based identification scheme based on discrete logarithms modulo a composite number. // Advances in Cryptology-EUROCRYPT'90. / Springer. 1991. P. 481–486.

[59] Hankerson D., Menezes A., Vanstone S. Guide to Elliptic Curve Cryptography. Springer-Verlag, New York, 2004. 358 p.

[60] ISO/IEC 9798-5:2009(E) «Information technology – Security techniques – Entity authentication – Part 5: Mechanisms using zero-knowledge techniques».

[61] Krawchuk H. HMAC: Keyed-Hashing for Message Authentication / H. Krawchuk, M. Bellare, R. Canetti. RFC 2104, 1997.

[62] Menezes A. Handbook of Applied Cryptography / A. Menezes, P. van Oorschot, and S. Vanstone. N.Y.:CRC Press, 1996.

[63] Pedersen T.P. Non-interactive and information-theoretic secure verifiable secret sharing. Proc. EUROCRYPT'91, Lect. Notes in Comput. Sci. 1992. Vol. 576. P. 129–140.

[64] Shamir A. How to share a secret. // Commun. ACM. New York City: ACM, 1979. Vol. 22, № 11. P. 612–613.