Ссылка на статью:
// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ. Электрон. журн. 2022, № 1, с. 72-79.

Поступила: 02.05.2022
Окончательный вариант: 05.05.2022

УДК 004.05

Сравнение с помощью тестов NIST зашифрованных последовательностей, полученных с помощью алгоритмов ГОСТ 34.12—2015 «Магма» и RSA

Смирнова А.В.

alinochka_gw@mail.ru
УлГУ, Ульяновск, Россия

Статья посвящена результатам массового статистического тестирования тестами NIST последовательностей открытого текста и последовательностей, зашифрованных алгоритмами шифрования ГОСТ 34.12—2015 «Магма» и RSA в режиме шифрования ECB

Ключевые слова*: криптография, случайные последовательности, статистическое тестирование, NIST STS, ГОСТ 34.12—2015 «Магма», RSA, режим шифрования ECB.*

Введение

Целью работы является исследование возможности определения алгоритма криптографического преобразования на основе тестов NIST [1]. Каждый алгоритм криптографического преобразования вырабатывает зашифрованные последовательности, по структуре которых с помощью тестов NIST их можно разбить на группы, причем признаком разбиения на группы будут служить результаты, полученные статистическим анализом зашифрованных последовательностей.

Каждый признак представляет собой множество результатов, параметров, комбинацию p-значений, полученных в результате тестов NIST. Список P-значений и соответствующих им тестов показан в таблице 3.

Задачами исследования является поиск признаков для идентификации группы, к которой принадлежит последовательность при рассмотрении распределений p–значений соответствующих тестов, а также результатов обучения нейронных сетей.

Предметом исследования являются следующие алгоритмы шифрования с использованием режима шифрования ECB: RSA и ГОСТ 34.12—2015 «Магма». В качестве тестируемых последовательностей используются файлы формата ZIP размера от 48 КБ до 6 МБ, содержащие текстовые файлы формата DOC.

Статистические тесты NIST применяются для анализа ГСЧ и ГСПЧ [6], а также они применялись для анализа алгоритмов шифрования, представленных на конкурс AES [2]. В [5] указаны некоторые методики статистического тестирования алгоритмов шифрования, в основном они заключаются в создании ГПСЧ на основе исследуемого алгоритма шифрования. В данной работе будет представлена другая методика.

Разные алгоритмы шифрования, а также генераторы псевдослучайных последовательностей имеют разные статистические характеристики, редко близкие по свойствам к случайным последовательностям, поэтому статистическое тестирование является одним из способов анализа криптографических свойств алгоритма шифрования.

1. Тесты NIST

Статистические тесты NIST возвращают одно или несколько P–значений, которые определяются как вероятность того, что идеальный генератор случайных чисел сгенерировал последовательность «менее случайную», чем исследуемая для статистики, определенной данным тестом [1].

Статистические тесты NIST представляют собой 15 тестов, целью каждого из которых является выявление определенного недостатка исследуемой последовательности, дающего судить о том, что данная последовательность не является случайной. Некоторые из тестов подразделяются на несколько подтестов и возвращают несколько P–значений.

2. Методика исследования

В качестве входных сообщений служили 630 файлов формата ZIP размера от 48 КБ до 6 МБ, содержащих текстовые файлы формата DOC. По всем имеющимся входным сообщениям, алгоритмам шифрования и ключам генерировались новые зашифрованные последовательности, затем проводилось статистическое тестирование тестами NIST всех входных и зашифрованных последовательностей. В работе исследовались следующие алгоритмы шифрования: ГОСТ 34.12—2015 «Магма» [3] (собственная реализация) и RSA (библиотека Openssl) в режиме простой замены (ECB) [4].

Количество битовых последовательностей на одно входное сообщение вычисляется по формуле (1).

$$n = 1 + n_k * n_c, \tag{1}$$

где n — количество битовых последовательностей, n_k — количество ключей, n_c — количество используемых алгоритмов шифрования,

Результирующие количественные характеристики представлены в таблице 1.

Таблица 1. Объем проведенного исследования

Числовая характеристика	Количество
Количество ключей алгоритма «Магма»	1000
Количество ключей алгоритма RSA	1000
Количество входных сообщений	630
Количество последовательностей на одно исследуемое сообщение	2 001
Количество последовательностей	1 260 630

После обработки полученных результатов тестирования происходило обучение нейронных сетей классификации, после чего был произведен анализ полученных нейронных сетей. Схема проведения исследования показана на рис. 1.

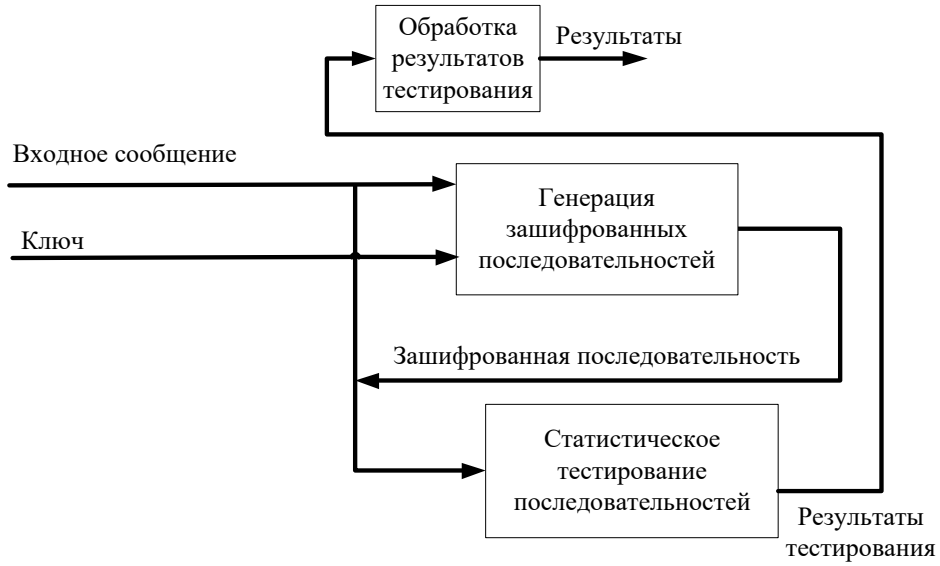

Рис. 1. Схема проведения исследования

3. Результаты статистического тестирования исследуемых последовательностей

Суть проведения исследования состоит в шифровании исходных последовательностей различными методами и последующей обработки с помощью полной системы тестов NIST. Каждый результат оценивается в виде вектора p-значений. Результаты представляются в виде диаграмм, по которым можно судить о принадлежности к той или иной группе классификации. Другими словами, любая группа характеризуется своим набором признаков из таблицы 2. Для демонстрации результатов было выбрано 10 тестов, возвращающих только одно P–значение.

Например, файлы формата zip отличаются тем, что во всех исследуемых последовательностях в частотном блочном тесте P–значения лежат в отрезке [0;0.1], поэтому, если у некоторой последовательности в этом тесте P–значение не лежит в этом отрезке, то можно предположить, что рассматривается зашифрованная последовательность.

Таблица 2. Соответствие тестов и столбцов таблицы результатов

Название теста	Столбцы таблицы, содержащие результат
Частотный побитовый тест	0
Частотный блочный тест	1
Тест на последовательность одинаковых бит	2
Тест на самую длинную последовательность единиц в блоке	3
Тест рангов бинарных матриц	4
Тест дискретного преобразования Фурье	5
Тест на совпадение неперекрывающихся шаблонов	6–153
Тест на совпадение перекрывающихся шаблонов	154
Универсальный статистический тест Маурера	155
Тест на линейную сложность	156
Проверка серий	157,158
Тест приблизительной энтропии	159
Тест кумулятивных сумм	160,161
Тест на произвольные отклонения	162–169
Вариант теста на произвольные отклонения	170–187

На рис. 2 изображены графики среднего значения и среднеквадратического отклонения для P–value по этим тестам. Обозначения тестов на горизонтальной оси приведены в таблице 2.

В таблице 3 указаны численные значения среднего значения и среднеквадратического отклонения P–значений по выбранным тестам в каждой группе.

Таблица 3. Среднее значение и среднеквадратическое отклонение P–value по выбранным тестам

Тест	Средние P–значения группы zip	Средние P–значения группы zip + «Магма»	Средние P–значения группы zip + «RSA»
Частотный побитовый тест	0.0404±0.1568	0.4677±0.2975	0.4128±0.3066
Частотный блочный тест	0	0.4698±0.3029	0.5065±0.3265

Тест на последовательность одинаковых бит	0.0143±0.0893	0.4675±0.2974	0.4188±0.306
Тест на самую длинную последовательность единиц в блоке	0.0598±0.1610	0.4583±0.2951	0.2598±0.2809
Тест рангов бинарных матриц	0.3595±0.3213	0.4171±0.3022	0.3779±0.3015
Тест дискретного преобразования Фурье	0.259±0.3015	0.4877±0.2916	0.0128±0.0864
Тест на совпадение перекрывающихся шаблонов	0.0416±0.1251	0.4922±0.2896	0.3205±0.2888
Универсальный статистических тест Маурера	$7\cdot10^{-5}$±0.0013	0.4383±0.304	0.462±0.2981
Тест на линейную сложность	0.4105±0.3022	0.4641±0.295	0.501±0.2881
Тест приблизительной энтропии	$3\cdot10^{-7}$±$6\cdot10^{-6}$	0.2554±0.3203	0.0045±0.0428

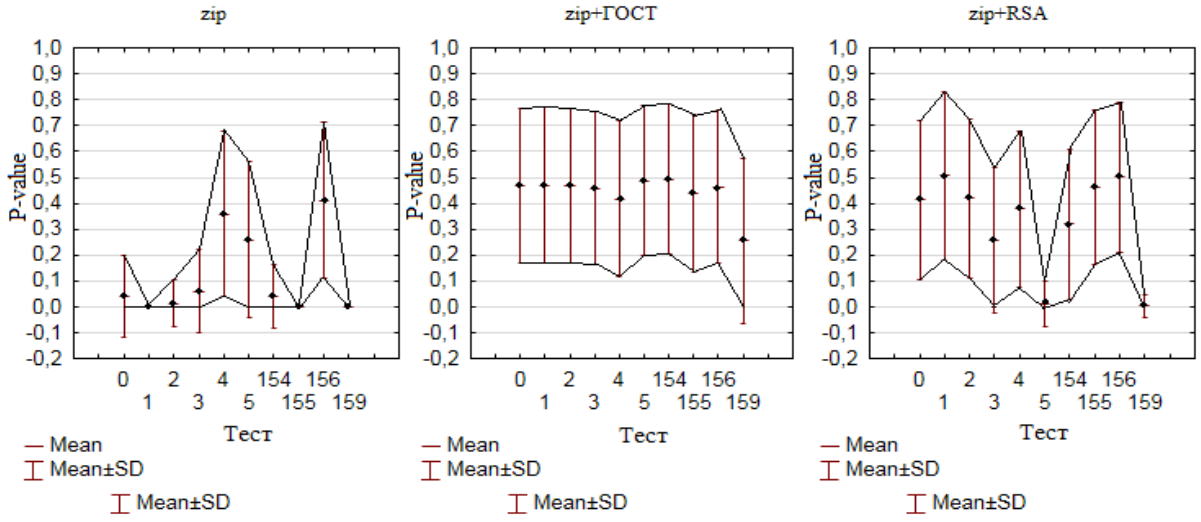

Рис. 2. Среднее значение (Mean) и среднеквадратическое отклонение (SD) P–value по выбранным тестам

Средние значения P–value у зашифрованных алгоритмом «Магма» файлов превышают все средние значения P–value исходных файлов в рассматриваемых тестах

Единственным тестом, в котором среднее значение P–value исходного файла превышает среднее значение P–value зашифрованных алгоритмом RSA файлов — это тест дискретного преобразования Фурье, что означает, что периодические свойства последовательности более ярко выражены в последовательностях, зашифрованных алгоритмом RSA, чем в файлах формата ZIP.

Кроме того, среднее значение P–value зашифрованных алгоритмом RSA файлов уступают средним значениям P–value зашифрованных алгоритмом «Магма» файлов более чем на 0.1 в следующих тестах: тест дискретного преобразования Фурье, тест приблизитель-

ной энтропии, тест на совпадение перекрывающихся шаблонов, тест на самую длинную последовательность единиц в блоке.

Режим шифрования ECB является наименее криптостойким, поэтому низкие средние p-значения, например, в тесте приблизительной энтропии, могут быть связаны с плохими статистическими характеристиками исходной последовательности.

На рис. 3 (см. Приложение) изображены графики среднего значения и среднеквадратического отклонения для P–value по всем тестам.

Заключение

По рис. 2 были определены отличительные признаки файлов формата zip и последовательностей, зашифрованных алгоритмами «Магма» и «RSA».

Для файлов zip признаками являются: с высокой вероятностью нахождение p-значения частотного побитового теста, частотного блочного теста, теста на последовательность одинаковых бит, теста на самую длинную последовательность единиц в блоке, теста на совпадение перекрывающихся шаблонов, универсального теста Маурера, теста приблизительной энтропии в промежутке [0,0.2], в тесте дискретного преобразования Фурье среднее p-значение равно 0.259±0.3015.

Укажем признаки для зашифрованных алгоритмом «Магма» файлов. В тесте на самую длинную последовательность единиц в блоке среднее p-значение равно 0.4583±0.2951, в тесте дискретного преобразования Фурье среднее p-значение равно 0.4877±0.2916, в тесте на совпадение перекрывающихся шаблонов среднее p-значение равно 0.4922±0.2896, в тесте приблизительной энтропии среднее p-значение равно 0.2554±0.3203.

Укажем признаки для зашифрованных алгоритмом «RSA» файлов. В тесте на самую длинную последовательность единиц в блоке среднее p-значение равно 0.2598±0.2809, в тесте дискретного преобразования Фурье среднее p-значение равно 0.0128±0.0864, в тесте на совпадение перекрывающихся шаблонов среднее p-значение равно 0.3205±0.2888, в тесте приблизительной энтропии среднее p-значение равно 0.0045±0.0428.

В будущем планируется более подробное рассмотрение результатов статистического тестирования, также планируется значительно увеличить количество входных последовательностей с добавлением файлов различных форматов, что необходимо для дальнейшего использования машинного обучения. Также, кроме данной гипотезы планируется проверить гипотезу о выявлении «слабых» ключей криптографического преобразования.

Список литературы

1. Bassham L., Rukhin A., Soto J., Nechvatal J., Smid M., Leigh S., Levenson M., Vangel M., Heckert N. and Banks D. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. Special Publication (NIST SP), National Institute of Standards and Technology. Gaithersburg, MD, 2010. Режим доступа:

https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906762 (дата обращения: 03.05.2022)
2. Soto J. and Bassham L. *Randomness Testing of the Advanced Encryption Standard Finalist Candidates*. NIST Interagency / Internal Report (NISTIR). National Institute of Standards and Technology. Gaithersburg, MD, 2000. https://doi.org/10.6028/NIST.IR.6483.
3. ГОСТ Р 34.12-2015. *Информационная технология (ИТ). Криптографическая защита информации. Блочные шифры*.
4. ГОСТ Р 34.13-2015. *Информационная технология (ИТ). Криптографическая защита информации. Режимы работы блочных шифров.*
5. Ключарёв П.Г. О статистическом тестировании блочных шифров // *Математика и математическое моделирование*. 2018, № 5, с. 35-56. https://doi.org/10.24108/mathm.0518.0000132.
6. Пикуза М.О., Михневич С.Ю. Тестирование аппаратного генератора случайных чисел при помощи набора статистических тестов NIST // *Доклады БГУИР*. 2021, т.19, вып. 4, с. 37-42. https://doi.org/10.35596/1729-7648-2021-19-4-37-42.

Comparison of sequences that are encrypted with GOST 34.12—2015 «Magma» and RSA using NIST statistical tests

Smirnova, A.V.

alinochka_gw@mail.ru

Ulyanovsk State University, Ulyanovsk, Russia

The paper presents comparison between results of mass statistical testing of plaintexts, GOST 34.12—2015 «Magma» ciphertexts and RSA ciphertexts in ECB mode of operations using NIST tests.

Keywords*: cryptography, random and pseudorandom number generators, statistical testing, NIST STS, RSA, GOST 34.12—2015 «Magma», ECB block mode of operations*

Приложение

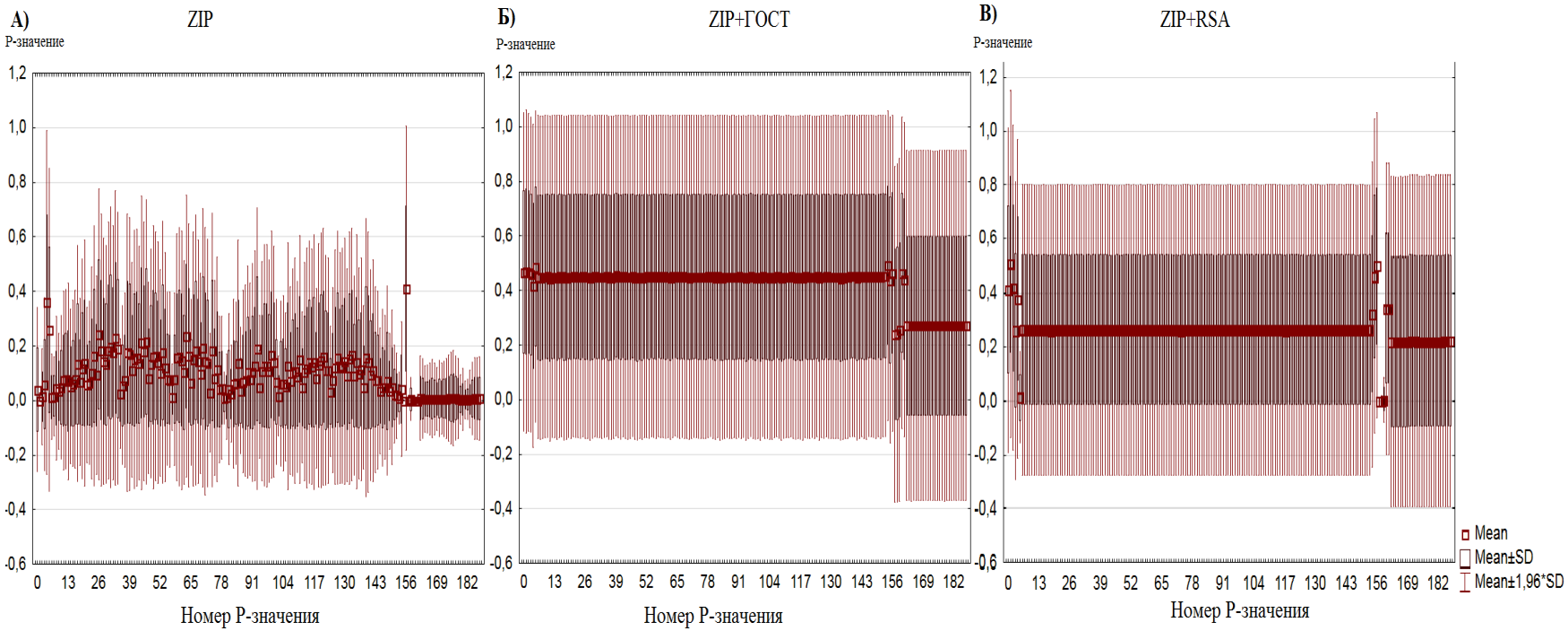

Рис. 3. Среднее значение (Mean) и среднеквадратическое отклонение (SD) Р–значений по всем тестам