СТАТИСТИЧЕСКИЙ АНАЛИЗ СОВРЕМЕННЫХ БЛОЧНЫХ ШИФРОВ*

А. И. Пестунов
Институт вычислительных технологий СО РАН, Новосибирск, Россия
e-mail: an24@ngs.ru

In this paper we present a statistical analysis of all block ciphers, which were AES (Advanced Encryption Standard) candidates. AES competition was organized by the US National Institute of Standards and Technology. With the help of the "Bookstack" test it is for the first time experimentally shown that a cipher text does not have a uniform distribution after 8 MARS rounds. It is experimentally ascertained that the FROG and LOKI97 ciphertexts can be distinguished from the uniform distribution after half of all rounds. We build a forecast that says that FROG and LOKI97 require 2^{78} and 2^{86} blocks respectively to distinguish a ciphertext from the uniform distribution after the full number of rounds.

Введение

Для безопасной передачи информации можно применять схемы с открытым ключом или схемы с закрытым (секретным) ключом. Использование открытого ключа очень удобно, когда передается небольшое количество информации, однако при больших ее объемах используется закрытый ключ. Базовым алгоритмом в схемах с закрытым ключом является блочный шифр, преобразующий текст по блокам фиксированной длины. Как правило, такие шифры состоят из трех процедур: развертывание секретного ключа, шифрование и дешифрирование. Вначале секретный ключ развертывается в массив подключей. Шифрование заключается в неоднократном преобразовании текста с помощью некоторой простой функции, зависящей от подключей и называемой раундом шифрования. Дешифрирование производится повторением раундов шифрования в обратном порядке. Чем больше раундов выполнено, тем труднее взломать шифр, однако тем медленней происходит шифрование, поэтому создатели шифров рекомендуют определенное число раундов, обеспечивающее как надежность, так и быстродействие алгоритма. В ряде шифров различные раунды отличаются друг от друга, что затрудняет их анализ.

Есть только один способ исследовать надежность шифра — это строить атаки на него, т. е. пытаться взломать. Построить атаку на шифр (произвести криптоанализ шифра) означает найти в нем какие-то недостатки. Главной целью являются разработка и реализация такого алгоритма, который за приемлемое время позволит найти секретный ключ или

*Работа выполнена при финансовой поддержке Президентской программы "Ведущие научные школы РФ" (грант № НШ-9886.2006.9) и Фонда содействия отечественной науке, грант "Лучшие аспиранты РАН".

все элементы массива подключей. Отметим, что знание секретного ключа эквивалентно знанию всего массива подключей в том смысле, что в обоих случаях шифр становится полностью детерминированным и расшифровать или зашифровать можно любое сообщение. Далеко не всегда удается построить атаку на R — полное число раундов шифра, поэтому интерес представляет даже криптоанализ $r < R$ их количества. Может быть, через годы или десятилетия эта атака распространится на полное число раундов. Важность такого рода исследований подтверждается тем, что в последнее время прошли несколько конкурсов на стандартизацию криптографических протоколов и, в частности, блочных шифров: американский AES (Advanced Encryption Standard) [1], японский CRYPTREC [2], европейский NESSIE [3]. Дизайнеры шифров и криптоаналитики со всего мира работали с целью найти лучшие алгоритмы. Все шифры — кандидаты конкурса AES, проводимого Национальным институтом стандартов и технологий США, должны были удовлетворять нескольким требованиям: размер блока равен 128 бит; секретный пользовательский ключ имеет длину 128, 192 или 256 бит; шифр должен быть надежным и быстрым. После нескольких лет исследований в 2001 году победителем и стандартом шифрования США был объявлен блочный шифр RIJNDAEL [4]. У других финалистов: RC6, TWOFISH, MARS и SERPENT, как и у победителя, не было найдено существенных недостатков, поэтому во многом благодаря красоте и простоте дизайна победил именно RIJNDAEL.

На данный момент не опубликовано атак, которые позволили бы взломать какой-либо из шифров финалистов AES с полным числом раундов. Более того, ни для одного из 15 кандидатов не было описано практически реализуемой атаки, действующей на полное число раундов. Шифр MARS [5] слабо изучен из-за его нестандартной структуры: он состоит из раундов четырех разных типов, по восемь каждого типа. В работе [6], описана атака на восемь раундов этого шифра (по два каждого типа), требующая 2^{25} блоков текста, 2^{29} байт памяти и не осуществимых на практике 2^{68} операций шифрования. Приведена также атака на 21 раунд, требующая 2^{232} операций шифрования, т. е. действующая незначительно быстрее полного перебора 256-битных ключей. Шифры LOKI97 [7] и FROG [8] не попали в финал, так как на них были описаны атаки, требующие 2^{56} блоков текста и того же порядка трудоемкости [9, 10], что существенно меньше, чем полный перебор ключей, но достаточно далеко до практической реализации.

Один из показателей надежности шифра — это неотличимость распределения шифртекста от равномерного, т. е. любой бит шифртекста должен принимать значение 0 или 1 с вероятностью $1/2$ независимо от других. Если удается показать, что распределение не равномерно, то имеют место статистические недостатки шифра (в англоязычной литературе поиск таких недостатков называется “distinguishing attack”).

В данной работе с помощью статистического теста “Стопка книг” [11] проведено исследование всех блочных шифров, участвовавших в конкурсе AES. Впервые экспериментально показано, что шифртекст после восьми раундов шифра MARS не подчиняется равномерному распределению, для этого потребовалось 2^{18} зашифрованных блоков. Экспериментально установлено, что шифртекст после половины всех раундов шифров LOKI97 и FROG не подчиняется равномерному распределению. Для этих двух шифров на основании полученных экспериментальных данных построен прогноз, позволяющий сказать, что с помощью 2^{78} и 2^{86} блоков шифртекст после полного числа раундов FROG и LOKI97 соответственно можно отличить от равномерного распределения.

1. Описание экспериментов

1.1. Статистический тест “Стопка книг”

Приведем краткое описание теста “Стопка книг” [11], с помощью которого мы будем статистически исследовать шифры. Тест предназначен для проверки гипотезы о том, что элементы выборки $Z = (z_1, z_2, \ldots, z_N)$ из алфавита $A = \{a_1, a_2, ..., a_S\}$ имеют равномерное распределение, т. е. они независимы и

$$\mathbf{P}(z_n = a_i) = 1/S;\ n = 1, \ldots, N;\ i = 1, \ldots, S.$$

Перед тестированием выборки в алфавите A фиксируется произвольный порядок, который меняется после анализа каждого выборочного элемента z_n следующим образом: буква z_n получает номер 1; номера тех букв, которые меньше номера этой буквы, увеличиваются на 1; у остальных букв номера не меняются. Формально эту процедуру можно описать так: пусть $\omega^n(a)$ — номер буквы $a \in A$ после анализа элементов $z_1, z_2, \ldots, z_{n-1}$, тогда

$$\omega^{n+1}(a) = \begin{cases} 1, & \text{если} z_n = a\,; \\ \omega^n(a) + 1, & \text{если} \omega^n(a) < \omega^n(z_n); \\ \omega^n(a), & \text{если} \omega^n(a) > \omega^n(z_n). \end{cases}$$

Такая конструкция похожа на стопку книг, если считать, что номер книги совпадает с ее положением в стопке. Книга извлекается из стопки и после чтения кладется наверх, ее номер становится первым. Книги, которые первоначально были над ней, сдвигаются вниз, а остальные остаются на месте. Множество всех номеров $\{1, \ldots, S\}$ разбивается на две непересекающиеся части — $A_1 = \{1, 2, \ldots, K\}$ и $A_2 = \{K + 1, \ldots, S\}$, где $K \in \{1, ..., S\}$ — параметр. Затем по выборке $(z_1, z_2, \ldots, z_N)$ подсчитывается ν_N — количество номеров $\omega^n(z_n)$, принадлежащих подмножеству A_1, т. е. количество попаданий букв в “верхнюю часть” “стопки книг”. $(N - \nu_N)$ — это количество попаданий в “нижнюю часть”. Далее вычисляется статистика:

$$x^2 = \frac{(\nu_N - NP_1)^2}{NP_1} + \frac{((N - \nu_N) - N(1 - P_1)^2)}{N(1 - P_1)},\ P_1 = |A_1|/S.$$

Если x^2 меньше критического уровня $\chi^2_{1,1-\alpha}$, то гипотеза о равномерности распределения элементов выборки принимается, иначе — отвергается. Величина $\chi^2_{1,1-\alpha}$— квантиль распределения хи-квадрат уровня значимости $(1 - \alpha)$ с одной степенью свободы. Если шифртекст не подчиняется равномерному распределению, то будем говорить, что имеют место “отклонения от случайности”, и чем больше x^2, тем отклонения от случайности “больше”. Разработанный нами алгоритм реализации теста “Стопка книг” на основе хеш-таблицы позволяет реализовать тест с помощью порядка $4 * K$ байт памяти и трудоемкостью $O(N)$.

1.2. Генерация выборки для тестирования

Как было замечено выше, требуется, чтобы шифртекст имел равномерное распределение, т. е. любой его бит должен принимать значения 0 или 1 независимо от других с вероятностью 1/2. Отсюда следует, что любая часть блока также должна иметь равномерное распределение. С помощью теста “Стопка книг” проверим выборку, составленную из таких частей блоков, и если элементы этой выборки не подчинятся равномерному распределению, то можно сделать вывод, что и шифртекст ему не подчинится.

Опишем то, как выборка составлялась. Все заявленные на конкурс AES шифры имели 128-битный блок, представим его в виде четырех 32-битных подблоков, занумерованных 0, 1, 2, 3, от старшего к младшему. Рассмотрим последовательность блоков $X_i^u, u \in \{0,1,2,3\}$, у которых подблок с номером u равен i, а остальные — нулевые, i пробегает значения $0, 1, ..., N-1$. Например, $X_7^0 = (7,0,0,0)$, $X_5^3 = (0,0,0,5)$. Обозначим за $y_i^{u,v}$ 32-битный подблок зашифрованного блока X_i^u с номером v. Например, зашифровав X_7^0, мы получим блок $(y_7^{0,0}, y_7^{0,1}, y_7^{0,2}, y_7^{0,3})$.

Зашифруем описанную последовательность X_i^u, $i = 0, 1, ..., N-1$, с помощью 100 случайных ключей и получим 100 выборок вида $(y_0^{u,v}, ..., y_{N-1}^{u,v})$ с фиксированными параметрами u и v, они предназначены для проверки, их элементы — это 32-битные подблоки. Будем брать $s = 8$, 16, 24 или 32 бита из каждого подблока, т. е. в разных случаях размер алфавита $S = 2^8, 2^{16}, 2^{24}$ или 2^{32}.

Для каждой из полученных выборок вычислим величины x^2 с заданным параметром K (размер верхней части "Стопка книг") и посчитаем значение $U_{99\,\%}$, означающее, сколько раз из ста эти величины превысили квантиль распределения хи-квадрат уровня значимости 0.99 ($\chi^2_{1,0.99} = 6.64$). Другими словами, сколько раз выборки не прошли тест.

Если распределение шифртекста для всех 100 ключей равномерное, то $P(x^2 > \chi^2_{1,0.99}) = 0.01$, и статистика

$$\chi^2(U_{99\,\%}) = \frac{(U_{99\,\%} - 100 * 0.01)^2}{100 * 0.01} + \frac{((100 - U_{99\,\%}) - 100(1 - 0.01))^2}{100(1 - 0.01)}$$

имеет распределение хи-квадрат. Отсюда заключаем, что $P(\chi^2(U_{99\,\%}) > \chi^2_{1,0.99999}) = 0.00001$ ($\chi^2_{1,0.99999} = 20$).

Непосредственной подстановкой легко убедиться, что если $U_{99\,\%} > 7$, то $\chi^2(U_{99\,\%}) > 20$, поэтому если $U_{99\,\%} > 7$, то с вероятностью 99.999 % можно утверждать, что распределение шифртекста не равномерно. В итоге наша задача сводилась к тому, чтобы найти параметры теста и размер выборки, обеспечивающие условие $U_{99\,\%} > 7$. Забегая вперед, отметим, что в экспериментах это значение было существенно больше 7.

2. Исследование шифра MARS

Перед использованием шифра MARS пользовательский ключ преобразуется в массив из сорока 32-битных подключей, затем для шифрования блока необходимо выполнить следующие действия.

1. Представить 128-битный блок в виде четырех 32-битных подблоков.
2. Сложить подблоки с первыми четырьмя подключами из массива.
3. Преобразовать блок с помощью:
   а) восьми FORWARD MIXING раундов;
   б) восьми FORWARD CORE раундов и двух подключей на каждом раунде;
   в) восьми BACKWARD CORE раундов и двух подключей на каждом раунде;
   г) восьми BACKWARD MIXING раундов.
4. Найти разность подблоков и последних четырех подключей из массива.

Если шифр состоит из однотипных раундов, то уменьшение их числа производится естественным образом: исследуется шифр, состоящий из одного раунда, двух, трех и т. д. Шифр MARS состоит из раундов различного типа, поэтому сократить их число можно разными способами. Мы рассмотрим несколько вариантов.

Таблица 1.

r	N	$U_{99\,\%}$	K	s	r	N	$U_{99\,\%}$	K	s
FORWARD MIXING раунды $u=3, v=2$					BACKWARD MIXING раунды $u=3, v=1$				
2	$\mathbf{2^6}$	100	2^6	8	2	$\mathbf{2^6}$	100	2^6	8
4	$\mathbf{2^6}$	100	2^6	8	4	$\mathbf{2^6}$	100	2^6	8
6	$\mathbf{2^{14}}$	67	2^{14}	24	6	$\mathbf{2^8}$	99	2^6	8
8	$\mathbf{2^{18}}$	43	2^{18}	32	8	$\mathbf{2^{14}}$	25	2^{18}	24
FORWARD CORE раунды $u=3, v=3$					BACKWARD CORE раунды $u=3, v=0$				
1	$\mathbf{2^6}$	100	2^6	8	1	$\mathbf{2^6}$	100	2^6	8
3	$\mathbf{2^6}$	100	2^6	8	3	$\mathbf{2^6}$	100	2^6	8
5	$\mathbf{2^{20}}$	17	2^{18}	32	5	$\mathbf{2^6}$	100	2^6	8
Симметричное сокращение $u=3, v=0$									
1+1+1+1	$\mathbf{2^6}$	100	2^6	8	2+2+2+2	$\mathbf{2^{18}}$	29	2^{18}	32

1. Шифрование происходит только с помощью:
   а) FORWARD MIXING раундов;
   б) BACKWARD MIXING раундов;
   в) FORWARD CORE раундов;
   г) BACKWARD CORE раундов.

2. Производится симметричное сокращение раундов (по одному или по два раунда каждого типа). Такой способ рассматривался в [6].

В результате экспериментов было замечено, что наибольшие отклонения от случайности получаются при $u = 3$. Значения v разные для разных модификаций шифра. В табл. 1 показаны результаты описанных экспериментов.

Из представленных результатов видно, что величина $U_{99\,\%}$ значительно больше 7, поэтому на основании выкладок, проведенных в подразд. 1.2, делаем вывод, что распределение шифртекста неравномерно с вероятностью 99.999 %.

3. Исследование шифров FROG и LOKI97

Описанные в разд. 2 эксперименты проводились для шифров FROG [8] и LOKI97 [7], которые состоят соответственно из восьми и 16 однотипных раундов. Для шифра FROG наибольшие отклонения от случайности достигались при $u = 2$ и $v = 2$ (табл. 2).

Для шифра LOKI97 параметры u и v были соответственно 1 и 2 для всех раундов, кроме восьмого, для него $u = 0$ и $v = 3$ (табл. 3).

Как и в случае с шифром MARS, $U_{99\,\%} > 7$, поэтому распределение шифртекста после указанного числа раундов шифров FROG и LOKI97 неравномерно.

В отличие от шифра MARS, использующего раунды различных типов, FROG и LOKI97 состоят из однотипных раундов, этот факт позволяет на основе экспериментальных ре-

Таблица 2.

r	N	$U_{99\,\%}$	K	s	r	N	$U_{99\,\%}$	K	s
1	2^9	100	2^8	16	3	2^{22}	20	2^{18}	32
2	2^{13}	22	2^{10}	16	4	2^{32}	17	2^{20}	32

Таблица 3.

r	N	$U_{99\,\%}$	K	s	r	N	$U_{99\,\%}$	K	s
1	2^3	100	2^3	6	5	2^{18}	52	2^{16}	32
2	2^5	100	2^3	6	6	2^{19}	66	2^{16}	32
3	2^8	100	2^6	8	7	2^{27}	31	2^{18}	32
4	2^{12}	100	2^{10}	16	8	2^{31}	22	2^{20}	32

зультатов спрогнозировать, сколько блоков необходимо для того, чтобы отличить от случайности шифртекст после большего числа раундов, когда эксперименты становятся невозможными из-за возрастающего количества требуемых блоков шифртекста.

Для обоих шифров наблюдается достаточно плавная, почти линейная зависимость величины $\log_2 N$ от числа раундов, поэтому приблизим эту зависимость крадратичной функцией

$$\log_2 \tilde{N}(r) = b_2 r^2 + b_1 r + b_0.$$

С помощью метода наименьших квадратов [12] найдем неизвестные коэффициенты b_0, b_1 и b_2. Пусть R^* — это то число раундов, которое удалось исследовать экспериментально ($R^* = 4$ для FROG и $R^* = 8$ для LOKI97), тогда неизвестные коэффициенты определятся из системы так называемых нормальных уравнений:

$$\sum_{i=0}^{R^*} \log_2 N_i = b_0(R^*+1) + b_1 \sum_{i=0}^{R^*} r_i + b_2 \sum_{i=0}^{R^*} r_i^2,$$

$$\sum_{i=0}^{R^*} \log_2 N_i r_i = b_0 \sum_{i=0}^{R^*} r_i + b_1 \sum_{i=0}^{R^*} r_i^2 + b_2 \sum_{i=0}^{R^*} r_i^3,$$

$$\sum_{i=0}^{R^*} \log_2 N_i r_i^2 = b_0 \sum_{i=0}^{R^*} r_i^2 + b_1 \sum_{i=0}^{R^*} r_i^3 + b_2 \sum_{i=0}^{R^*} r_i^4.$$

После решения системы для шифра FROG зависимость размера выборки от числа раундов примет вид $\log_2 \tilde{N}(r) = 0.5r^2 + 5.7r + 0.8$ (табл. 4).

Для шифра LOKI97 получена зависимость $\log_2 \tilde{N}(r) = 0.2r^2 + 2.5r - 0.3$ (табл. 5).

Таблица 4.

	Эксперименты					Прогноз			
r_i	0	1	2	3	4	5	6	7	8
N_i	2^0	2^9	2^{13}	2^{22}	2^{32}	2^{42}	2^{53}	2^{65}	2^{78}

Таблица 5.

	Эксперименты									Прогноз			
r_i	0	1	2	3	4	5	6	7	8	10	12	14	16
N_i	2^0	2^3	2^5	2^8	2^{12}	2^{18}	2^{19}	2^{27}	2^{31}	2^{43}	2^{56}	2^{70}	2^{86}

Таблица 6.

Шифр	r	R	N	$U_{99\,\%}$	K	s	u	v
E2	3	12	2^{20}	100	2^{18}	32	1	0
DFC	3	8	2^{20}	100	2^{18}	32	0	1
CAST256	2	12	2^{18}	68	2^{18}	32	0	1
RIJNDAEL	2	10,12,14	2^{18}	40	2^{18}	32	2	2
CRYPTON	3	12	2^{5}	100	2^{5}	8	0	1
SERPENT	3	32	2^{31}	31	2^{20}	32	0	3
SAFER+	2	8,12,16	2^{30}	77	2^{20}	32	1	2
DEAL	2	6,8,8	2^{6}	100	2^{5}	8	2	2
MAGENTA	2	6,6,8	2^{6}	100	2^{5}	8	0	1
HPC	1	8	2^{10}	100	2^{10}	16	2	2
RC6	4	20	2^{27}	47	2^{20}	32	0	2
TWOFISH	3	16	2^{20}	52	2^{18}	32	2	0

Таким образом, можно заключить, что распределение шифртекста шифра FROG после полного числа раундов, равного восьми, можно отличить от равномерного с помощью статистического теста “Стопка книг”, имея порядка 2^{78} блоков шифртекста, а LOKI97 — после полного числа раундов, равного 16, — порядка 2^{86}.

4. Исследование остальных кандидатов конкурса AES

Остальные 12 кандидатов конкурса AES также были исследованы с помощью теста “Стопка книг”. В табл. 6 приведены число раундов, после которого шифртекст можно отличить от случайности (r), полное число раундов (R), размер выборки, на который фиксируются отклонения (N) и параметры тестирования, введенные в разд. 2. Для шифров RIJNDAEL, MAGENTA, SAFER+ и DEAL число раундов зависит от длины секретного пользовательского ключа, поэтому в таблице даны все возможные значения. Например, для RIJNDAEL при 128-битном ключе нужно выполнить 10 раундов, 192-битном — 12 и 256-битном — 14.

Список литературы

[1] Advanced Encryption Algorithm (AES) Development Effort // 1997–2000. http://csrc.nist.gov/encryption/aes

[2] CRYPTREC project // 2000–2002. http://www.ipa.go.jp/security/enc/CRYPTREC

[3] New European Schemes for Signatures, Integrity, and Encryption // Deliverables of the NESSIE Project, 2003. http://www.cosic.esat.kuleuven.ac.be/nessie

[4] Daemen J., Rijmen V. The Rijndael Block Cipher // AES submission. 1999. http://csrc.nist.gov/encryption/aes/rijndael/Rijndael.pdf

[5] Burwick C. et. al. MARS — a candidate cipher for AES // AES submission. 1999. http://www.research.ibm.com/security/mars.pdf

[6] Kelsey J., Schneier B. MARS Attacks! Preliminary Cryptanalysis of Reduced-Round MARS Variants // Proc. Third AES Candidate Conf., 2000. http://www.schneier.com/paper-mars-attacks.pdf

[7] Brown L., Pieprzyk J. Introducing the New LOKI97 Block Cipher // AES submission. 1998. http://www.adfa.oz.au/ lpb/research/loki97/loki97spec.ps

[8] Georgoudis D., Leroux D., Chaves B. The FROG Encryption Algorithm // AES submission. 1998. http://csrc.nist.gov/encryption/aes/round1/conf1/frog-slides.pdf

[9] Knudsen L., Rijmen V. Weaknesses in LOKI97 // Proc. Second AES Candidate Conf., 1999. http://www.adfa.oz.au/ lpb/research/loki97/knudsen99.pdf

[10] Wagner D., Ferguson N., Schneier B. Cryptanalysis of FROG // Proc. Second AES Candidate Conf., 1999. http://www.schneier.com/paper-frog.pdf

[11] Рябко Б.Я., Пестунов А.И. "Стопка книг" как новый статистический тест для случайных чисел // Пробл. передачи информации. 2004. Т. 40, вып. 1. С. 73–78.

[12] Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа: Руководство для экономистов: Пер. с нем. / Предисл. В.М. Ивановой. М.: Финансы и статистика, 1983. 302 с.

Поступила в редакцию 13 декабря 2006 г.,
в переработанном виде — 18 января 2007 г.