СТАТИСТИЧЕСКИЕ АНАЛОГИ В КРИПТОАНАЛИЗЕ СИММЕТРИЧНЫХ БЛОЧНЫХ ШИФРОВ

Г. П. Агибалов, И. А. Панкратова

Томский государственный университет
Томск, *Россия*
E-mail: agibalov@isc.tsu.ru, pank@isc.tsu.ru

Для функции шифрования вводится понятие статистического аналога (СА) как булева уравнения, выполняющегося с некоторой вероятностью, связывающего переменные ключа, открытого и шифрованного текстов и обладающего свойством: функция в нем (после замены переменной шифртекста функцией шифрования) статистически не зависит от переменной открытого текста. Изложены три алгоритма криптоанализа произвольного симметричного блочного шифра путем решения нелинейного СА его функции шифрования методом максимального правдоподобия. Один алгоритм исходит из условно разделимого СА, два других – из СА с малым линеаризационным множеством. Одним из последних можно найти 34 бита ключа DES, в то время как известный алгоритм M. Matsui при тех же условиях находит только 26 бит.

Ключевые слова: симметричный блочный шифр, статистический аналог, криптоанализ.

ВВЕДЕНИЕ

Содержание этого доклада является частью статьи авторов [1], посвященной статистическим аналогам дискретных функций и их применению в криптоанализе симметричных блочных шифров. Для функции шифрования $F(x, k)$ статистический аналог определяется как приближенное соотношение (*approximate expression*, или *statistical relation*), введенное Mitsuru Matsui [2, 3], и является булевым уравнением $\varphi(x, y, k) = 0$, не обязательно линейным, которое выполняется с некоторой вероятностью p, связывает переменные символов открытого текста – x, его шифртекста – y и ключа – k и, в отличие от приближенного соотношения M. Matsui, обладает свойством статистической независимости ассоциированной с ним булевой функции $\varphi(x, F(x, k), k)$ от переменных в x. Это свойство гарантирует сохранение вероятности аналога после подстановки в него открытого текста и соответствующего шифртекста, чего может не быть для приближенного соотношения.

В [1] сформулированы конструктивные тесты статистической независимости булевой функции от подмножества ее аргументов. Так, функция $f(x, y)$, где x, y – переменные со значениями в $(Z_2)^n$ и $(Z_2)^m$ соответственно, статистически не зависит от переменных в x, если и только если функция $f(x, y) \oplus (u, x)$ уравновешена для любого ненулевого вектора $u \in (Z_2)^n$. Доказана сохраняемость вероятности статистического аналога при любом фиксировании в нем символов соответствующих открытого и шифрованного текстов. Доказано, что сумма статистических аналогов является также статистическим аналогом, и приведена формула для вероятности последнего. Для суперпозиции дискретных функций $h=g(f(x\oplus k), y)$ показано, что уравнение $\psi(f(x\oplus k), y) = 0$, где ψ – функция статистического аналога для g, является статистическим аналогом для h с той же вероятностью. Изложены методы построения линейных и нелинейных статистических аналогов для функций блоков замены, раундовых функций и многораундовых шифров с аддитивным раундовым ключом и алгоритмы криптоанализа произвольных симметричных блочных шифров путем решения систем линейных и нелинейных статистических аналогов функций шифрования методом максимального правдоподобия. Изложение иллюстрировано примерами из криптоанализа DES. Показано также, что один из предложенных алгоритмов позволяет на основе пары нелинейных статистических аналогов 16-раундового DES, построенных M. Matsui, найти 34 бита ключа DES, в то время как алгоритм самого M. Matsui [3] на основе тех же двух приближенных соотношений получает только 26 из этих бит.

Кроме того, в [4] доказано, что если x, y, z – наборы булевых переменных и функция $f(x, y)$ статистически не зависит от переменных в x, то последнее верно и для любой функции $g(f(x, y), z)$, но может быть неверным для функции $g(f_1(x, y), \ldots, f_s(x, y, z)$, где $s \geq 2$ и каждая из функций $f_1(x, y), \ldots, f_s(x, y)$ статистически не зависит от x.

Ниже, после введения необходимых определений, приводятся из [1] только алгоритмы криптоанализа произвольных симметричных блочных шифров, построенные на основе нелинейных статистических аналогов функций шифрования.

ОПРЕДЕЛЕНИЯ

Для любой булевой функции f и для любого подмножества U ее аргументов будем говорить, что *f статистически не зависит* от переменных множества U, если для любой ее подфункции f', полученной фиксированием значений всех переменных в U, имеет место $\Pr[f' = 0] = \Pr[f = 0]$, где для булевой функции g от s переменных, имеющей в своем векторе значений ровно $w_0(g)$ символов 0, $\Pr[g = 0] = w_0(g)/2^s$.

Рассмотрим произвольную функцию F: $X \times K \to Y$, где $X = (Z_2)^n$, $K = (Z_2)^m$, $Y = (Z_2)^r$ для некоторых натуральных n, r и целого $m \geq 0$. В частности, F может быть функцией одного раунда итеративного блочного шифра, и тогда X и Y суть множества блоков соответственно на входе и выходе раунда, а K – множество раундовых ключей. Ею может быть и функция симметричного шифрования открытых текстов из X в шифртексты из Y на ключах из K. При $m = 0$ функция F рассматривается как отображение F: $X \to Y$. В этом случае она может быть функцией, например, бесключевого блока замены. Следующие определения предполагают $m > 0$.

Статистическим аналогом (далее СА) функции F называется всякое уравнение $\varphi(x, y, k) = 0$, в котором $x = x_1x_2 \ldots x_n$, $y = y_1y_2 \ldots y_r$, $k = k_1k_2 \ldots k_m$ – переменные (булевы векторы) со значениями в X, Y, K соответственно, связанные соотношением $y = F(x, k)$, и φ: $X \times Y \times K \to Z_2$ – булева функция от $n + m + r$ переменных, существенно зависящая хотя бы от одной переменной в каждом из наборов x, y и k, такая, что функция $\varphi_F(x, k) = \varphi(x, F(x, k), k)$, называемая *ассоциированной* с этим СА, статистически не зависит от переменных в x. Число $p = \Pr[\varphi_F = 0]$ называется *вероятностью* данного СА. Говорят также, что он *выполняется с вероятностью p* и имеет *эффективность* $\varepsilon = |p - 1/2|$. СА называют *эффективным*, если $p \neq 1/2$, или, что то же самое, $\varepsilon > 0$. Функция φ в нем называется *функцией* самого аналога, который, в свою очередь, именуется как СА, *заданный* этой функцией.

Эти определения легко переписываются на случай $m = 0$, а именно: опускаются все вхождения символа k и требование статистической независимости φ_F от x. Таким образом, в этом случае фактически имеем дело с функцией $F(x)$, с ее СА $\varphi(x, y) = 0$, где $\varphi(x, y)$ – любая булева функция от $n + r$ переменных, и с его вероятностью $p = \Pr[\varphi_F(x) = 0]$, где $\varphi_F(x) = \varphi(x, F(x))$.

В случае $m > 0$ статистическая независимость функции $\varphi_F(x, k)$ от x придает заданному функцией φ СА функции F следующее важное свойство: фиксирование в уравнении СА для F любого значения x и того значения y, в которое F преобразует это x при равновероятно выбранном k, не изменяет вероятности выполнения этого уравнения [1].

Заметим, что свойство статистической независимости ассоциированной функции $\varphi_F(x, k)$ от x, обусловливающее наше понятие статистического аналога функции F, существенно отличает его от других понятий того же предназначения, известных под названиями *approximate expression*, *statistical relation* и т. п. и не предполагающих данного свойства. В его же отсутствие может непредсказуемо измениться вероятность используемого в криптоанализе approximate expression (statistical relation и т. п.) после подстановки в него открытого текста и соответствующего шифртекста, что сделает практически неэффективным алгоритм криптоанализа, основываемый обычно на решении системы вероятностных уравнений методом максимального правдоподобия (МП).

СА с нелинейной функцией φ называется далее *нелинейным* или *НСА*.

КРИПТОАНАЛИЗ НА ОСНОВЕ НСА

Чтобы подчеркнуть единородство этого метода и линейного криптоанализа, будем называть его *нелинейным криптоанализом*, видя в словах «линейный» и «нелинейный» не противоположность, привносимую частицей «не», но, прежде всего, единство их корня. Нелинейный криптоанализ, как и линейный, направлен на частичное раскрытие ключа шифра и может быть атакой как с известным открытым текстом, так и с выбором оного. Но в любом случае для реализации нелинейного криптоанализа предполагается наличие некоторого эффективного нелинейного статистического аналога функции шифрования. Возможны разные алгоритмы нелинейного криптоанализа, основанные на методе МП и использующие разные свойства заданного НСА. Здесь мы представим три таких алгоритма из [1]: один предполагает в НСА свойство условной разделимости, два других – свойство малости линеаризационного множества. Первые два алгоритма являются атаками с известным открытым текстом, третий – атакой с выбором открытого текста.

Пусть для рассматриваемого симметричного шифра имеется НСА $\varphi(x, y, k) = 0$ функции шифрования $F(x, k)$, выполняемый с некоторой вероятностью p, и $0 < p < 1$.

КРИПТОАНАЛИЗ НА ОСНОВЕ УСЛОВНО-РАЗДЕЛИМОГО НСА

Представим заданный НСА как $\varphi'(x, y, k') = (1, k'')$, где k' и k'' – наборы некоторых переменных в k, не имеющие общих переменных, $(1, k'')$ – сумма всех тех переменных в k, если таковые есть, которые входят только в линейные слагаемые полинома Жегалкина (АНФ) функции φ (по ним φ линейная), и $\varphi'(x, y, k')$ есть сумма остальных слагаемых в полиноме. В отсутствие переменных в k, по которым функция $\varphi(x, y, k)$ линейная, считаем $(1, k'') = 0$ и $\varphi'(x, y, k') = \varphi(x, y, k)$. Пусть также x' есть набор всех тех переменных в x, которые являются существенными аргументами функции φ (входят в ее АНФ явно).

Будем называть НСА $\varphi = 0$ *разделимым* (по переменным), если ассоциированная с ним функция $\varphi_F(x, k) = \varphi(x, F(x, k), k)$ статистически не зависит от переменных в наборе $x'k'$ и $|k''| > 1$. Если, кроме того, число переменных в k' сравнительно мало (в пределах трёх-четырёх десятков – с позиции производительности современных компьютеров), то данный НСА называется *условно разделимым*. Важность этого понятия очевидна: в случае статистической независимости φ_F от $x'k'$ любое фиксирование открытого текста x, соответствующего шифртекста y и значений переменных в k' в уравнении $\varphi(x, y, k) = 0$, приводит его к линейному уравнению с неизвестными в k'', выполнимому с той же вероятностью, что и $\varphi = 0$.

Теперь неизвестные значения переменных в k' и один бит информации о k'' могут быть найдены следующим алгоритмом, где N – количество известных открытых текстов.

1. Для каждого из возможных значений $k'^{(j)}$ набора k' ($j = 1, 2, ..., 2^{|k'|}$) и для каждой пары известных открытого текста $x^{(i)}$ и его шифртекста $y^{(i)}$ ($i = 1, 2, ..., N$) определяется $\varphi'(x^{(i)}, y^{(i)}, k'^{(j)})$.

2. За значение k' берём то $k'^{(j)}$, при котором $\varphi'(x^{(i)}, y^{(i)}, k'^{(j)})$ со всевозможными парами $(x^{(i)}, y^{(i)})$ принимает некоторое значение $d \in \{0,1\}$ чаще (другого) при $p > 1/2$ и реже при $p < 1/2$.

3. Полагаем $(1, k'') = d$, тем самым находим еще один бит информации о ключе k.

КРИПТОАНАЛИЗ НА ОСНОВЕ НСА С МАЛЫМ ЛИНЕАРИЗАЦИОННЫМ МНОЖЕСТВОМ

Атака с известным открытым текстом. Подставив в заданный НСА $\varphi(x, y, k) = 0$ вместо x и y соответственно известные открытые тексты $x^{(i)} \in X$ и их криптограммы $y^{(i)} \in Y$ для $i = 1, 2, ..., N$, получим систему булевых уравнений для компонент неизвестного ключа k, а именно:

$$\varphi_i(k) = 0,\ i = 1, 2, ..., N, \tag{1}$$

где $\varphi_i(k) = \varphi(x^{(i)}, y^{(i)}, k)$ для всех $i \in \{1, 2, ..., N\}$. Каждое уравнение в системе (1) вероятностное и выполняется с вероятностью p.

Следуя [5], назовем подмножество переменных в k *линеаризационным*, если при фиксации любых их значений каждое уравнение в системе (1) превращается в линейное (линеаризуется). Зафиксируем некоторое (лучше – наименьшей мощности, или кратчайшее) линеаризационное множество L переменных в системе (1). Для каждого набора $L^{(j)}$ значений переменных в L возьмем подфункцию $\varphi_i^{(j)}(k')$ функции $\varphi_i(k)$, полученную подстановкой вместо переменных в L их значений в наборе $L^{(j)}$. Здесь $j = 1, 2, ..., s = 2^{|L|}$. Ввиду свойства линеаризационного множества функция $\varphi_i^{(j)}(k')$ является аффинной. Пусть $\varphi_i^{(j)}(k') = (c_i^{(j)}, k') \oplus d_i^{(j)}$. Таким образом, получаем систему линейных уравнений

$$(c_i^{(j)}, k') = d_i^{(j)},\ i = 1, 2, ..., N; j = 1, 2, ..., s, \tag{2}$$

где каждое уравнение выполняется с вероятностью $q = s^{-1}p$. Эта система представляет собой объединение s подсистем $E_1, ..., E_s$, где E_j для любого $j \in \{1,2, ..., s\}$ состоит из уравнений в (2) для $i = 1, 2, ..., N$. Каждая подсистема E_j решается алгоритмом 1 в [2], а именно: полагаем $t_j = N - \sum_{i=1}^{N} d_i^{(j)}$, вычисляем

1) $d^{(j)} = 0$, если $t_j > N/2$ и $q > 1/2$ или $t_j \le N/2$ и $q < 1/2$;

2) $d^{(j)} = 1$, если $t_j \le N/2$ и $q > 1/2$ или $t_j > N/2$ и $q < 1/2$,

и записываем детерминированную систему уравнений

$$(c_i^{(j)}, k') = d^{(j)},\ i = 1, 2, ..., N.$$

Если эта система совместна, то ее решение относительно k' вместе с набором $L^{(j)}$ является результатом криптоанализа – предполагаемым ключом шифра. Это надо понимать так, что если последняя система совместна при нескольких значениях $j \in \{1, 2, ..., s\}$, то результат криптоанализа будет неоднозначным, что, естественно, возможно при недостаточном количестве N использованных пар (открытый текст, шифртекст).

Ясно, что данный алгоритм реально выполним лишь тогда, когда линеаризационное множество L достаточно мало. Непосредственно проверяется, что уравнения (4) и (5) в [3], являющиеся НСА для 16-раундового DES, этим свойством обладают: множества $L_3 = \{K_1[18], K_1[19], ..., K_1[23], K_{16}[42], K_{16}[43], ..., K_{16}[47]\}$ и $L_4 = \{K_1[42], K_1[43], ..., K_1[47], K_{16}[18], K_{16}[19], ..., K_{16}[23]\}$ являются линеаризационными в системе (1) для этих НСА соответственно. Таким образом, применив данный алгоритм с этими НСА, можно получить 22 бита ключа DES, а именно: 12 бит в L_3 и 10 бит из правой части (4) в [3] или 12 бит в L_4 и 10 бит из правой части (5) в [3]. Если же применить алгоритм сначала, скажем, с первым НСА, а затем – со вторым, то можно получить 44 бита раундовых ключей DES или, с учетом расписания ключей, 34 бита исходного ключа, в то время как алгоритм M. Matsui в [3] позволяет получить от тех же самых двух НСА только 26 из этих бит.

Атака с выбором открытого текста. Мощность линеаризационного множества переменных в системе (1) зависит как от вида φ, так и от того, какие именно пары открытых текстов $x^{(i)} \in X$ и их криптограмм $y^{(i)} \in Y$ для каждого $i = 1, 2, ..., N$ подставлены в НСА $\varphi(x, y, k) = 0$ с целью получения этой системы. Если выбрать открытые тексты такими, при которых система (1), полученная подстановкой их и соответствующих шифртекстов в уравнение $\varphi(x, y, k) = 0$, будет иметь линеаризационное множество L наименьшей мощности (или близкой к нему), и использовать это L в последнем алгоритме криптоанализа, то можно достичь максимальной (или близкой к ней) скорости выполнения данного алгоритма (при заданной функции φ). Это и будет атака с выбором открытого текста. Дальнейшее ее ускорение возможно на пути выбора более подходящего НСА. Впрочем, последнее замечание относится и к атаке с известным открытым текстом.

ЛИТЕРАТУРА

1. *Агибалов, Г. П.* Элементы теории статистических аналогов дискретных функций с применением в криптоанализе итеративных блочных шифров / Г. П. Агибалов, И. А. Панкратова // Прикладная дискретная математика. 2010. № 3. С. 51–68.

2. *Matsui, M.* Linear Cryptanalysis Method for DES Cipher / M. Matsui // LNCS. 1993. Vol. 765. P. 386–397.

3. *Matsui, M.* The First Experimental Cryptanalysis of the Data Encryption Standard / M. Matsui // LNCS. 1994. Vol. 839. P. 1–11.

4. *Колчева, О. Л.* О статистической независимости суперпозиции булевых функций / О. Л. Колчева, И. А. Панкратова // Прикладная дискретная математика. Приложение. 2011. № 4. С. 11–12.

5. *Агибалов, Г. П.* Логические уравнения в криптоанализе генераторов ключевого потока / Г. П. Агибалов // Вестн. Томского гос. ун-та. Приложение. 2003. № 6. С. 31–41.