МАТЕМАТИЧЕСКИЕ МЕТОДЫ КРИПТОГРАФИИ

DOI 10.17223/20710410/9/5

УДК 519.7

ЭЛЕМЕНТЫ ТЕОРИИ СТАТИСТИЧЕСКИХ АНАЛОГОВ ДИСКРЕТНЫХ ФУНКЦИЙ С ПРИМЕНЕНИЕМ В КРИПТОАНАЛИЗЕ ИТЕРАТИВНЫХ БЛОЧНЫХ ШИФРОВ[1]

Г. П. Агибалов, И. А. Панкратова

Томский государственный университет, г. Томск, Россия

E-mail: agibalov@isc.tsu.ru, pank@isc.tsu.ru

Вводится понятие статистической независимости булевой функции от подмножества аргументов. На его основе определяется понятие статистического аналога дискретной функции как булева уравнения, выполняемого с некоторой вероятностью, и изучаются его свойства. Формулируются конструктивные тесты статистической независимости. Излагаются методы построения линейных статистических аналогов функций итеративного блочного шифрования с аддитивным раундовым ключом и некоторые алгоритмы криптоанализа симметричных шифров, основанные на решении систем линейных и нелинейных статистических аналогов методом максимального правдоподобия. Приводимые определения, методы и алгоритмы иллюстрируются на примере DES. В частности, показано, что одним из алгоритмов криптоанализа, предложенных в статье, можно найти 34 бита ключа 16-раундового DES, используя пару известных статистических аналогов, на базе которых алгоритм M. Matsui доставляет только 26 из этих бит. Статья может служить учебно-методическим пособием по теме в заголовке, в том числе по линейному криптоанализу.

Ключевые слова: *статистическая независимость, статистические аналоги функций, итеративные блочные шифры, криптоанализ, линейный криптоанализ, нелинейный криптоанализ, DES.*

Введение

Известно, что в криптоанализе двоичных симметричных шифров значительную роль играют системы булевых уравнений и методы их решения [1]. Системой уравнений переменные биты открытого текста и соответствующего шифртекста связываются с неизвестными битами ключа в соответствии с алгоритмом шифрования. Путём решения этой системы с известными битами открытого текста и шифртекста как раз и достигается цель криптоанализа — находятся (все или некоторые) биты ключа. Известно, однако, что решение произвольной системы нелинейных уравнений (пусть даже степени 2) имеет экспоненциальную сложность. Есть много приёмов «упрощения» систем уравнений, благодаря которым система, не поддающаяся никакому методу, после упрощения нередко становится решаемой некоторым методом за приемлемое время. Один из таких приёмов, восходящий к [2, 3], заключается в замене заданной системы уравнений E системой её так называемых приближённых соотношений (approximate

[1]Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (гос. контракт № П1010).

expressions), где каждое соотношение e яляется булевым уравнением, связывающим переменные системы E и выполняющимся с некоторой вероятностью $p \neq 1/2$. Чем более простыми (в некотором смысле) выбраны приближённые соотношения для системы E (например, линейными), тем легче решается система из них с подставленными открытыми и шифртекстами, а чем больше количество соотношений и выше эффективность каждого соотношения в ней, выражающаяся для e разностью $|p - 1/2|$, тем выше результативность полученного решения, т. е. вероятность того, что корень системы приближённых соотношений будет корнем системы E. Именно так устроен, например, линейный криптоанализ двоичных симметричных шифров, в котором вместо системы уравнений шифрования применяется система её линейных приближённых соотношений с ненулевыми эффективностями, решаемая при известных значениях бит открытого текста и шифртекста алгоритмом полиномиальной сложности. Для достижения достаточно высокой результативности такого решения обычно требуется иметь много различных открытых текстов и соответствующих шифртекстов, чтобы получить достаточно большое число независимых уравнений из системы приближённых соотношений.

В данной работе, написанной, главным образом, с целью уточнения, формализации и дальнейшего развития понятийного аппарата линейного криптоанализа Mitsuru Matsui [2], излагаются элементы теории статистических аналогов, выступающих в криптоанализе в роли приближённых соотношений M. Matsui, не обязательно линейных. Они вводятся для функций шифрования как булевы уравнения, которые выполняются с некоторой вероятностью, связывают переменные символов открытого текста, его шифртекста и ключа и обладают свойством статистической независимости ассоциированных с ними булевых функций от переменных символов открытого текста. Последнее свойство существенным образом отличает статистический аналог от приближённого соотношения и гарантирует сохранение вероятности аналога после подстановки в него открытого текста и соответствующего шифртекта, чего может не быть для приближённого соотношения. Формулируются конструктивные тесты статистической независимости булевой функции от подмножества её аргументов. Доказывается сохраняемость вероятности статистического аналога при любом фиксировании в нём символов соответствующих открытого и шифрованного текстов. Для суперпозиции двух дискретных функций определяется суперпозиция одной из них (внутренней) и статистического аналога другой (внешней) и показывается, что в случае аддитивности внутренней функции полученная суперпозиция является функцией статистического аналога для первой суперпозиции с вероятностью статистического аналога её внешней функции. Излагаются методы построения линейных статистических аналогов для функций блоков замены, раундовых функций и многораундовых шифров с аддитивным раундовым ключом и алгоритмы криптоанализа итеративных блочных шифров путём решения систем линейных и нелинейных статистических аналогов функций шифрования методом максимального правдоподобия. Изложение иллюстрируется примерами из криптоанализа DES. Показано, в частности, что один из предложенных здесь алгоритмов криптоанализа 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 приближённых соотношений получает только 26 из этих бит. Работа может быть рекомендована в качестве учебно-методического пособия по рассматриваемой теме.

1. Статистическая независимость

Для любой булевой функции f и для любого подмножества U её аргументов будем говорить, что f *статистически не зависит* от переменных множества U, если для любой её 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 своём векторе значений ровно $w_0(g)$ символов 0, $\Pr[g = 0] = w_0(g)2^{-s}$.

Пусть далее $\oplus$ есть сложение в $\mathbb{Z}_2$, т. е. по mod 2. Считаем, что в применении к булевым векторам эта операция выполняется покомпонентно.

Утверждение 1. Функция $f(x, k) = g(x \oplus k)$, где x, k — переменные со значениями в $(\mathbb{Z}_2)^n$, статистически не зависит от переменных в x.

Доказательство. В самом деле, $w_0(f) = 2^n w_0(g)$ и $g(x \oplus k)$ при любом фиксированном $x = a$ пробегает всё множество значений функции g. Таким образом, $\Pr[f(x, k) = 0] = w_0(f)/2^{2n} = w_0(g)/2^n$ и $\Pr[f(a, k) = 0] = w_0(g(a \oplus k))/2^n = w_0(g)/2^n$. ■

Через (a, b) будем обозначать скалярное произведение булевых векторов a и b.

Утверждение 2. Функция $f(x, y)$, где x, y — переменные со значениями в $(\mathbb{Z}_2)^n$ и $(\mathbb{Z}_2)^m$ соответственно, статистически не зависит от переменных в x, если и только если функция $f(x, y) \oplus (u, x)$ уравновешена для любого ненулевого вектора $u \in (\mathbb{Z}_2)^n$.

Доказательство. Обозначим через $w_1(f)$ вес функции f (количество единиц в её векторе значений). Непосредственно проверяется, что f статистически не зависит от переменных в x, если и только если $w_1(f(a, y)) = w_1(f)/2^n$ для любого вектора $a \in (\mathbb{Z}_2)^n$.

Необходимость. Разложим функцию $f(x, y) \oplus (u, x)$ по всем переменным в x; коэффициенты этого разложения имеют вид $f_a(y) = f(a, y) \oplus (u, a)$ для всевозможных $a \in (\mathbb{Z}_2)^n$. Если $(u, a) = 0$ (а это условие при фиксированном ненулевом u выполняется ровно для половины всех a), то $w_1(f_a) = w_1(f(a, y)) = w_1(f)/2^n$. Если же $(u, a) = 1$, то $w_1(f_a) = 2^m - w_1(f(a, y)) = 2^m - w_1(f)/2^n$. Известно, что вес функции равен сумме весов коэффициентов её разложения; запишем:

$$w_1(f(x, y) \oplus (u, x)) = 2^{n-1} w_1(f)/2^n + 2^{n-1}(2^m - w_1(f)/2^n) = 2^{n+m-1},$$

что и доказывает уравновешенность функции $f(x, y) \oplus (u, x)$.

Достаточность. Докажем сначала, что $w_1(f(a, y)) = w_1(f)/2^n$ для нулевого вектора a. Снова запишем вес функции $f(x, y) \oplus (u, x)$ как сумму весов коэффициентов разложения и учтём уравновешенность этой функции:

$$w_1(f(x, y) \oplus (u, x)) = 2^{n+m-1} = \sum_{a \in (\mathbb{Z}_2)^n} w_1(f(a, y) \oplus (u, a)) =$$

$$= \sum_{a, (u,a)=0} w_1(f(a, y)) + \sum_{a, (u,a)=1} \left(2^m - w_1(f(a, y))\right),$$

откуда $\sum\limits_{a, (u,a)=0} w_1(f(a, y)) = \sum\limits_{a, (u,a)=1} w_1(f(a, y))$. Просуммируем обе части последнего равенства по всем $u \neq 0$:

$$\sum_{u \neq 0} \sum_{\substack{a, \\ (u,a)=0}} w_1(f(a, y)) = \sum_{u \neq 0} \sum_{\substack{a, \\ (u,a)=1}} w_1(f(a, y)).$$

Заметим, что при любом фиксированном $a \neq 0$ и всевозможных $u \neq 0$ равенство $(u, a) = 1$ выполняется 2^{n-1} раз, а равенство $(u, a) = 0$ верно в остальных $(2^{n-1} - 1)$

случаях. При $a = 0$ всегда $(u, a) = 0$. Поэтому получим

$$(2^n - 1)w_1(f(0, y)) + (2^{n-1} - 1)\sum_{a\neq 0} w_1(f(a, y)) = 2^{n-1}\sum_{a\neq 0} w_1(f(a, y)),$$

откуда

$$\sum_{a\in(\mathbb{Z}_2)^n} w_1(f(a, y)) = 2^n w_1(f(0, y)$$

и $w_1(f(0, y)) = w_1(f)/2^n$.

Для случая $a \neq 0$ рассмотрим функцию $g(x, y) = f(x \oplus a, y)$. Ясно, что $f(a, y) = g(0, y)$; кроме того, функция $g(x, y)\oplus(u, x)$ уравновешена в случае уравновешенности $f(x, y) \oplus (u, x)$, так как

$$w_1(f(x, y) \oplus (u, x)) = \sum_{x,y}(f(x, y) \oplus (u, x)) =$$
$$= \sum_{x,y}(f(x \oplus a, y) \oplus (u, x \oplus a)) = \sum_{x,y}(g(x, y) \oplus (u, x) \oplus (u, a)).$$

Последняя сумма здесь (в зависимости от значения (u, a)) есть вес функции $g(x, y) \oplus (u, x)$ или её отрицания, что для уравновешенной функции одно и то же. По доказанному выше $w_1(g(0, y)) = w_1(g)/2^n$, т. е. $w_1(f(a, y)) = w_1(f)/2^n$. ■

С использованием преобразования Уолша — Адамара (см., например, [4]) тест может быть переформулирован следующим (более конструктивным) образом: функция $f(x, y)$ статистически не зависит от переменных в x, если и только если для любого ненулевого вектора $u \in (\mathbb{Z}_2)^n$ имеет место равенство $\hat{f}(u, 0) = 0$, где $\hat{f}$ — преобразование Уолша — Адамара функции f.

2. Понятие статистического аналога

2.1. Основные определения

Рассмотрим произвольную функцию $F : X \times K \to Y$, где $X = (\mathbb{Z}_2)^n$, $K = (\mathbb{Z}_2)^m$, $Y = (\mathbb{Z}_2)^r$ для некоторых натуральных n, r и целого $m \geqslant 0$. 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 $m \geqslant 1$.

Статистическим аналогом (*СА*) функции F называется всякое уравнение $\varphi(x, y, k) = 0$, в котором $x = x_1x_2...x_n$, $y = y_1y_2...y_r$, $k = k_1k_2...k_m$ — переменные (булевы векторы) со значениями в X, Y, K соответственно, связанные соотношением $y = F(x, k)$, и $\varphi : X \times Y \times K \to \mathbb{Z}_2$ — 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 нём 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 её 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 придаёт 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. Строго говоря, верно следующее

Утверждение 3. Пусть СА $\varphi(x,y,k)=0$ функции $F(x,k)$ имеет вероятность p. Пусть также $x^{(i)}$ — произвольное значение переменной x и $y^{(i)}=F(x^{(i)},k)$ для некоторого k, выбранного в K случайно с вероятностью 2^{-m}. Тогда $\Pr[\varphi(x^{(i)},y^{(i)},k)=0]=p$.

Доказательство. В самом деле, $\Pr[\varphi(x^{(i)},y^{(i)},k)=0]=\Pr[\varphi_F(x^{(i)},k)=0]=$ $=\Pr[\varphi'_F(k)=0]=\Pr[\varphi_F(x,k)=0]=p$. ■

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 шифра 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.

Класс функции φ некоторого СА (в некоторой классификации булевых функций) называется также *классом* этого СА. В частности, статистический аналог называется *линейным* (*ЛСА*), если его функция φ линейная, т. е. если $\varphi(x,y,k)=(a,x)\oplus$ $\oplus(b,y)\oplus(c,k)$ для некоторых констант $a\in X\setminus\{0\}$, $b\in Y\setminus\{0\}$, $c\in K\setminus\{0\}$. Нередко ЛСА $(a,x)\oplus(b,y)\oplus(c,k)=0$ записывается как $(a,x)\oplus(b,y)=(c,k)$. В случае $m=0$ он имеет вид $(a,x)\oplus(b,y)=0$.

СА функции F, принадлежащий некоторому классу C, называется *оптимальным* (в классе C), если его эффективность наибольшая среди эффективностей всех СА этой функции, входящих в C. Таким образом, можно говорить, например, об оптимальных ЛСА для данной функции F.

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

Пример 1. Пусть $n=m=r$ и $y=F(x,k)=x\oplus k$. Тогда для любого ЛСА $(a,x)\oplus(b,y)=(c,k)$ для F с некоторой вероятностью p выполняется уравнение $(a,x)\oplus$ $\oplus(b,x\oplus k)=(c,k)$, или, что то же самое, $(a\oplus b,x)=(b\oplus c,k)$. Возможны два случая.

1) $a=b$. В этом случае имеем $(b\oplus c,k)=0$, $p=\Pr[(b\oplus c,k)=0]$, и следовательно, если $b=c$, то $p=1$ и $\varepsilon=1/2$, а если $b\neq c$, то $p=1/2$ и $\varepsilon=0$.

2) $a\neq b$. В этом случае получаем уравнение $(a\oplus b,x)=(b\oplus c,k)$, которое при каждом k выполняется для половины возможных значений x, поэтому $p=1/2$ и $\varepsilon=0$.

Таким образом, СА вида $(a,x\oplus y\oplus k)=0$, и только они являются эффективными ЛСА для функции $x\oplus k$.

Пример 2. Пусть $n=3$, $m=r=1$, $y=F(x,k)=x_1k_1\oplus x_1x_3\oplus x_2x_3$, $a=100$, $b=c=1$. Тогда ЛСА $(a,x)\oplus(b,y)\oplus(c,k)=0$ для F имеет вероятность p, с которой выполняется уравнение $x_1\oplus x_1k_1\oplus x_1x_3\oplus x_2x_3\oplus k_1=0$. Левая часть последнего обращается в 0 на шести из шестнадцати возможных наборов значений переменных в ней, поэтому $p=3/8$ и $\varepsilon=1/8$.

2.2. Статистический аналог суперпозиции функций

Многие функции шифрования строятся как суперпозиции других, более простых функций, в связи с чем возникает задача построения функции статистического аналога суперпозиции из её компонент и их статистических аналогов. Здесь мы рассмотрим эту задачу в ситуации, когда функция F представлена суперпозицией других функций как $F(x,k) = G(H(x,k))$, где $H : X \times K \to Z$, $G : Z \to Y$ и $Z = (\mathbb{Z}_2)^l$ для некоторого $l \geqslant 1$. Примером такого представления F может служить суперпозиция $S(x \oplus k)$ функции замены (S) и суммы $(\oplus)$ заменяемого информационного блока (x) и раундового ключа (k) в итеративных блочных шифрах с аддитивным раундовым ключом [5], в частности в DES.

Пусть уравнение $\psi(z,y) = 0$ является СА функции $G(z)$ и $p = \Pr[\psi(z, G(z)) = 0]$ — его вероятность. Построим функцию $\varphi(x,y,k) = \psi(H(x,k),y)$. Будем иметь $\varphi_F(x,k) = \varphi(x, F(x,k), k) = \psi(H(x,k), F(x,k)) = \psi(H(x,k), G(H(x,k)))$. Спрашивается, является ли уравнение $\varphi(x,y,k) = 0$ статистическим аналогом для F, или, равносильно, зависит ли статистически функция $\varphi_F(x,k)$ от переменных в наборе x. В каждом конкретном случае ответ на этот вопрос можно получить с помощью теста статистической независимости либо проверив непосредственно выполнение равенства $\Pr[\varphi'_F(k) = 0] = \Pr[\varphi_F(x,k) = 0]$, где $\varphi'_F(k) = \psi(H'(k), G(H'(k)))$ и $H'(k)$ — произвольная подфункция функции $H(x,k)$, полученная фиксированием под знаком последней значений всех переменных в x.

В случае $X = K = Z$ и $F(x,k) = G(x \oplus k)$, т. е. когда $H(x,k) = x \oplus k$, функцию $F(x,k)$ называют *функцией с аддитивным параметром* — k. В этом случае $\varphi(x,y,k) = \psi(x \oplus k, y)$.

Утверждение 4. Для функции F с аддитивным параметром функция $\varphi_F(x,k)$ статистически не зависит от x.

Доказательство. Утверждение справедливо в силу утверждения 1 при $f = \varphi_F$ и $g(x \oplus k) = \psi(x \oplus k, G(x \oplus k))$. ■

Следствие 1. Для функции $F(x,k)$ с аддитивным параметром уравнение $\psi(x \oplus k, y) = 0$ является статистическим аналогом.

Утверждение 5. Вероятность статистического аналога $\psi(x \oplus k, y) = 0$ функции $F(x,k)$ с аддитивным параметром равна p.

Доказательство. В самом деле, $\Pr[\psi(x \oplus k, G(x \oplus k)) = 0] = 2^{-n}|\{x \oplus k \in Z : \psi(x \oplus k, G(x \oplus k)) = 0\}| = 2^{-n}|\{z \in Z : \psi(z, G(z)) = 0\}| = \Pr[\psi(z, G(z)) = 0] = p$. ■

2.3. Сложение статистических аналогов

Покажем, что множество всех статистических аналогов одной и той же функции замкнуто относительно сложения (по частям) в поле $\mathbb{Z}_2$ различных и независимых СА, и приведём формулу для вероятности суммы таких СА. В этой связи индукцией по натуральному s докажем следующую лемму, известную по [2] как Piling-up Lemma.

Лемма 1. Для s независимых случайных переменных X_i с $\Pr[X_i = 0] = p_i$ и $\Pr[X_i = 1] = 1 - p_i$ для $i = 1, 2, ..., s$ вероятность $\Pr[X_1 \oplus X_2 \oplus ... \oplus X_s = 0]$ вычисляется как $q_s = 1/2 + 2^{s-1}\Pi_{i=1}^{s}(p_i - 1/2)$.

Доказательство. В самом деле, при $s = 1$ это очевидно. Предположим, что это верно при некотором $s \geqslant 1$, т. е. $\Pr[X_1 \oplus X_2 \oplus ... \oplus X_s = 0] = q_s$, и докажем, что $\Pr[X_1 \oplus ... \oplus X_s \oplus X_{s+1} = 0] = q_{s+1}$. Сумма по mod 2 равна 0 тогда и только тогда, когда оба её слагаемых равны одновременно 0 или 1, поэтому $\Pr[X_1 \oplus ... \oplus X_s \oplus X_{s+1} = 0] =$

$= q_s p_{s+1} + (1-q_s)(1-p_{s+1})$. Положим здесь $q = q_s - 1/2$ и $p = p_{s+1} - 1/2$. Тогда $\Pr[X_1 \oplus ... \oplus X_s \oplus X_{s+1} = 0] = (q+1/2)(p+1/2) + (1/2-q)(1/2-p) = 1/2 + 2pq =$ $= 1/2 + 2(p_{s+1} - 1/2)(q_s - 1/2) = 1/2 + 2 \cdot 2^{s-1} \Pi_{i=1}^{s}(p_i - 1/2)(p_{s+1} - 1/2) = q_{s+1}$. ■

Утверждение 6. Пусть $\varphi_1(x,y,k) = 0$ и $\varphi_2(x,y,k) = 0$ — различные и независимые СА функции F и $\varphi = \varphi_1 \oplus \varphi_2$. Тогда $\varphi(x,y,k) = 0$ есть также СА функции F.

Доказательство. В случае $m = |k| = 0$ утверждение очевидно. Пусть $m > 0$. Требуется доказать статистическую независимость φ_F от x. Пусть p_1 и p_2 — вероятности заданных в условии СА соответственно. Тогда $\Pr[\varphi'_{1F} = 0] = \Pr[\varphi_{1F} = 0] = p_1$, $\Pr[\varphi'_{2F} = 0] = \Pr[\varphi_{2F} = 0] = p_2$ и в силу леммы 1 $\Pr[\varphi'_F = 0] = \Pr[\varphi'_{1F} \oplus \varphi'_{2F} = 0] =$ $=1/2 + 2(p_1 - 1/2)(p_2 - 1/2) = \Pr[\varphi_{1F} \oplus \varphi_{2F} = 0] = \Pr[\varphi_F = 0]$. ■

Таким образом, доказано, что сумма любых s различных и независимых статистических аналогов некоторой функции с вероятностями $p_1, p_2, ..., p_s$ соответственно действительно является СА этой функции, и его вероятность вычисляется по формуле для q_s в лемме 1. Его эффективность ε, как видно, не превосходит эффективности любого из слагаемых. В частности, если $p_i = 1/2$ хотя бы для одного $i \in \{1, ..., s\}$, то $\varepsilon = 0$.

Все приводимые далее статистические аналоги, как линейные, так и нелинейные, для функций в DES заимствованы из литературы по *линейному* криптоанализу, где они подаются под названием *linear* approximate equations (relations, expressions), см., например, [2, 3].

3. Линейные статистические аналоги для DES

На примере DES рассмотрим методы построения эффективных ЛСА для блоков замены, раундовых функций и функций шифрования многораундовых итеративных блочных симметричных шифров.

3.1. ЛСА блоков замены DES

Блоки замены в DES по традиции будем называть S-блоками. Рассмотрим сначала функцию любого S-блока $S_i : (\mathbb{Z}_2)^6 \to (\mathbb{Z}_2)^4, i \in \{1, 2, ..., 8\}$, и произвольный её ЛСА $(a,x) \oplus (b,y) = 0$. Здесь $a \in (\mathbb{Z}_2)^6, b \in (\mathbb{Z}_2)^4$, x и y — переменные со значениями в $(\mathbb{Z}_2)^6$ и $(\mathbb{Z}_2)^4$ соответственно и $y = S_i(x)$. Определим $N_i(a,b) = |\{x \in (\mathbb{Z}_2)^6 : (a,x) = (b, S_i(x))\}|$. Например, $N_1(011011, 0100) = 22$, $N_5(010000, 1111) = 12$. По определению, $2^{-6}N_i(a,b)$ есть вероятность, с которой выполняется равенство $(a,x) \oplus (b, S_i(x)) = 0$ при равновероятном выборе $x \in (\mathbb{Z}_2)^6$, поэтому $\Pr[(a,x) \oplus (b,y) = 0] = 2^{-6}N_i(a,b)$. Так, $\Pr[(011011, x) \oplus (0100, S_1(x)) = 0] = 22/64 = 11/32$, $\Pr[(010000, x) \oplus (1111, S_5(x)) = 0] =$ $= 12/64 = 3/16$. Вычислив $2^{-6}N_5(a,b)$ для всех пар ab в $(\mathbb{Z}_2)^6 \times (\mathbb{Z}_2)^4$, т. е. вероятности всевозможных ЛСА функции S_5, можно убедиться, что ЛСА $(010000, x) \oplus (1111, y) = 0$ является оптимальным (в классе линейных СА) для S_5 (с вероятностью $p = 3/16$ и эффективностью $\varepsilon = 5/16$). В таблице приведены найденные таким образом оптимальные линейные статистические аналоги для всех восьми S-блоков DES, их вероятности p и эффективности ε. Из неё видно, что ЛСА, указанный для S_5, имеет наибольшую эффективность среди эффективностей ЛСА всех S-блоков DES.

Оптимальные ЛСА для S-блоков DES

Номер S-блока	a	b	p	ε
1	010000	1111	7/32	9/32
2	100010	1011	1/4	1/4
3	100010	1111	1/4	1/4
4	100010	1111	1/4	1/4
	101000	1111	1/4	1/4
	101011	0110	1/4	1/4
	101011	1001	1/4	1/4
5	010000	1111	3/16	5/16
6	010000	0111	9/32	7/32
	100010	1011	9/32	7/32
7	111011	0100	7/32	9/32
8	010000	1111	1/4	1/4
	100010	1110	1/4	1/4

3.2. ЛСА раундовой функции DES

Пусть для произвольного булева вектора $v = v_{t-1}v_{t-2}...v_0$ и для любых различных $i_1, i_2, ..., i_s$ в $\{0, 1, ..., t-1\}$ символ $v(i_1, i_2, ..., i_s)$ обозначает $(i_1, i_2, ..., i_s)$-проекцию вектора v, т. е. $v(i_1, i_2, ..., i_s) = v_{i_1}v_{i_2}...v_{i_s}$, а символ $v[i_1, i_2, ..., i_s]$ — сумму по mod 2 всех компонент этой проекции, т. е. $v[i_1, i_2, ..., i_s] = v_{i_1} \oplus v_{i_2} \oplus ... \oplus v_{i_s}$.

Функция одного раунда в DES является отображением $F : X \times K \to Y$, где $X = (\mathbb{Z}_2)^n$, $K = (\mathbb{Z}_2)^m$, $Y = (\mathbb{Z}_2)^r$ для $n = r = 64$, $m = 48$, и для любых $k \in K$ и $x = x_L x_R \in X$, где $|x_L| = |x_R|$, определяется равенством $F(x, k) = y$, в котором $y = y_L y_R \in Y$, $y_L = x_R$, $y_R = x_L \oplus f(x_R, k)$ для некоторой функции $f : (\mathbb{Z}_2)^{32} \times (\mathbb{Z}_2)^{48} \to (\mathbb{Z}_2)^{32}$. Последняя является суперпозицией элементарных операций над булевыми векторами (расширение, перестановка, сложение по mod 2) и функций S-блоков, такой, что для любого номера S-блока $i = 1, 2, ..., 8$ существуют $i_1, i_2, ..., i_6$ и l_1, l_2, l_3, l_4 в $\{0, 1, ..., 31\}$, а также $j_1, j_2, ..., j_6$ в $\{0, 1, ..., 47\}$, для которых

$$f(x_R, k)(l_1, l_2, l_3, l_4) = S_i(x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6)). \tag{1}$$

Здесь и далее в изложении, относящемся к DES, предполагается, что компоненты векторов $x \in X$, $y \in Y$ и $k \in K$ занумерованы справа налево целыми числами, начиная с 0. В этом предположении верно, в частности, $x_R(t) = x(t)$ для $0 \leqslant t \leqslant 31$.

Равенство (1) означает, что функция $F(x, k) = f(x_R, k)(l_1, l_2, l_3, l_4)$ получена суперпозицией функции $G = S_i$ и функции $x \oplus k = x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6)$ и, таким образом, является функцией с аддитивным параметром.

Возьмём любой ЛСА i-го S-блока $(a, u) \oplus (b, v) = 0$ с некоторыми вероятностью p_i и эффективностью ε_i. В нём $a \in (\mathbb{Z}_2)^6$, $b \in (\mathbb{Z}_2)^4$, u и v — переменные со значениями в $(\mathbb{Z}_2)^6$ и $(\mathbb{Z}_2)^4$ соответственно и $v = S_i(u)$. Подставив в него сначала $S_i(u)$ вместо v, а затем $x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6)$ вместо u и формулу $f(x_R, k)(l_1, l_2, l_3, l_4)$ вместо равной ей по (1) подформулы $S_i(x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6))$, получим уравнение $(a, x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6)) = (b, f(x_R, k)(l_1, l_2, l_3, l_4))$. Обозначив в последнем $f(x_R, k)$ как y_f (переменный булев вектор длиной 32), придём к следующему линейному уравнению:

$$(a, x_R(i_1, i_2, ..., i_6)) \oplus (b, y_f(l_1, l_2, l_3, l_4)) = (a, k(j_1, j_2, ..., j_6)). \tag{2}$$

В нём функция, равная сумме его левой и правой частей, построена как суперпозиция функции ψ $(= (a,u) \oplus (b,v))$ статистического аналога для G $(= S_i)$ и функции $x \oplus k$ $(= x_R(i_1, i_2, ..., i_6) \oplus k(j_1, j_2, ..., j_6))$. Следовательно, по следствию 1, уравнение (2) является статистическим аналогом и тем самым ЛСА функции f. По утверждению 5 его вероятность равна p_i, а эффективность — ε_i.

Например, если взяты 5-й S-блок и его ЛСА из таблицы, т. е. если $i = 5$ и $a = 010000$, $b = 1111$, то $(i_1, i_2, ..., i_6) = (16, 15, ..., 11)$, $(j_1, j_2, ..., j_6) = (23, 22, ..., 18)$, $(l_1, l_2, l_3, l_4) = (24, 18, 7, 29)$ и ЛСА (2) для f имеет вид

$$x_R[15] \oplus y_f[7, 18, 24, 29] = k[22]. \tag{3}$$

Его вероятность и эффективность равны соответственно $3/16$ и $5/16$ (те же, что у взятого ЛСА 5-го S-блока).

Аналогичным образом строятся ЛСА для f по ЛСА остальных S-блоков, в том числе и по не приведённым в таблице. Так, ЛСА для f, построенный по ЛСА $(011011, u) \oplus (0100, v) = 0$ для S-блока S_1, есть

$$x_R[27, 28, 30, 31] \oplus y_f[15] = k[42, 43, 45, 46] \tag{4}$$

и выполняется с вероятностью $11/32$. Следующие три ЛСА функции f построены этим же методом по другим ЛСА блоков S_1 и S_5:

$$x_R[29] \oplus y_f[15] = k[44]; \tag{5}$$

$$x_R[15] \oplus y_f[7, 18, 24] = k[22]; \tag{6}$$

$$x_R[12, 16] \oplus y_f[7, 18, 24] = k[19, 23]. \tag{7}$$

Их вероятности равны $15/32$, $21/32$, $1/4$ соответственно.

Кроме того, новые ЛСА функции f можно строить как суммы уже построенных для неё ЛСА (утверждение 6). Так, сумма (3) и (4) является ЛСА $x_R[15, 27, 28, 30, 31] \oplus \oplus y_f[7, 15, 18, 24, 29] = k[22, 42, 43, 45, 46]$ для f с вероятностью $1/2+2(3/16-1/2)(11/32-1/2) \approx 0{,}6$. Поскольку эффективность суммы статистических аналогов не выше эффективности слагаемых, а эффективность ЛСА (3) наибольшая среди эффективностей ЛСА всех S-блоков, то построенные так новые ЛСА будут иметь эффективность не выше $5/16$ — эффективности ЛСА (3).

Как следует из приведённых построений, ЛСА для f в общем виде представляется уравнением

$$x_R[i_1, i_2, ..., i_{s_1}] \oplus y_f[l_1, l_2, ..., l_{s_2}] = k[j_1, j_2, ..., i_{s_3}], \tag{8}$$

выполняемым с некоторой вероятностью p. В нём s_1, s_2, s_3 — некоторые натуральные числа, x_R, y_f, k — переменные со значениями в $(\mathbb{Z}_2)^{32}$, $(\mathbb{Z}_2)^{32}$, $(\mathbb{Z}_2)^{48}$ соответственно, $0 \leqslant i_t \leqslant 31$, $0 \leqslant l_t \leqslant 31$ и $0 \leqslant j_t \leqslant 47$ для всех подходящих t.

На i-м раунде DES, $i \geqslant 1$, пара 32-битных векторов $L_{i-1}R_{i-1}$ преобразуется по раундовому ключу K_i в пару 32-битных векторов L_iR_i по правилам

$$L_i = R_{i-1},\ R_i = L_{i-1} \oplus f(R_{i-1}, K_i). \tag{9}$$

Положив в (8) $x_R = R_{i-1}$, $k = K_i$ и ввиду (9) $y_f = f(x_R, k) = f(R_{i-1}, K_i) = R_i \oplus L_{i-1}$, получим ЛСА для i-го раунда DES

$$R_{i-1}[i_1, i_2, ..., i_{s_1}] \oplus L_{i-1}[l_1, l_2, ..., l_{s_2}] \oplus R_i[l_1, l_2, ..., l_{s_2}] = K_i[j_1, j_2, ..., i_{s_3}]$$

с вероятностью p. Полезно помнить, что в нём ввиду (9) $R_{i-1} = L_i$.

Применяя данный метод, построим, в качестве примера, некоторые ЛСА для первых пяти раундов DES. Потом они пригодятся нам в построении ЛСА для многораундовых DES.

Так, положив в (3) $x_R = R_0$, $k = K_1$ и $y_f = f(x_R, k) = f(R_0, K_1) = L_0 \oplus R_1$, получим следующий ЛСА 1-го раунда DES, выполненный, как и (3), с вероятностью $3/16$:

$$R_0[15] \oplus L_0[7, 18, 24, 29] \oplus R_1[7, 18, 24, 29] = K_1[22]. \tag{10}$$

Положив в (3) $x_R = R_1 = L_2$, $k = K_2$ и $y_f = f(x_R, k) = f(R_1, K_2) = L_1 \oplus R_2$, получим ЛСА 2-го раунда DES также с вероятностью $3/16$:

$$L_1[7, 18, 24, 29] \oplus L_2[15] \oplus R_2[7, 18, 24, 29] = K_2[22]. \tag{11}$$

Положив же в (3) $x_R = R_2 = L_3$, $k = K_3$ и $y_f = f(x_R, k) = f(R_2, K_3) = L_2 \oplus R_3$, получим ЛСА 3-го раунда DES, выполненный опять же с вероятностью $3/16$:

$$L_3[15] \oplus L_2[7, 18, 24, 29] \oplus R_3[7, 18, 24, 29] = K_3[22]. \tag{12}$$

При $x_R = R_3$, $k = K_4$ и $y_f = f(x_R, k) = f(R_3, K_4) = L_3 \oplus R_4$ в (3) имеем ЛСА 4-го раунда DES с вероятностью $3/16$:

$$R_3[15] \oplus L_3[7, 18, 24, 29] \oplus R_4[7, 18, 24, 29] = K_4[22]. \tag{13}$$

Аналогичными заменами из (4) получаются ещё один ЛСА для 1-го раунда DES, но уже с вероятностью $11/32$

$$R_0[27, 28, 30, 31] \oplus L_0[15] \oplus R_1[15] = K_1[42, 43, 45, 46] \tag{14}$$

и ЛСА для 5-го раунда DES с вероятностью $11/32$

$$L_5[27, 28, 30, 31] \oplus L_4[15] \oplus R_5[15] = K_5[42, 43, 45, 46]. \tag{15}$$

3.3. ЛСА многораундовых DES

Линейные статистические аналоги для многораундовых DES строятся путём суммирования нескольких ЛСА одиночных раундов DES. Так, сумма (10) и (12), равная

$$R_0[15] \oplus L_0[7, 18, 24, 29] \oplus L_3[15] \oplus R_3[7, 18, 24, 29] = K_1[22] \oplus K_3[22],$$

является ЛСА 3-раундового DES с вероятностью $1/2 + 2(3/16 - 1/2)(3/16 - 1/2) = 0{,}70$, а сумма (11), (13), (14) и (15), равная

$$\begin{aligned} &L_0[15] \oplus R_0[7, 18, 24, 27, 28, 29, 30, 31] \oplus L_5[7, 18, 24, 27, 28, 29, 30, 31] \oplus R_5[15] = \\ &= K_1[42, 43, 45, 46] \oplus K_2[22] \oplus K_4[22] \oplus K_5[42, 43, 45, 46], \end{aligned} \tag{16}$$

— ЛСА 5-раундового DES с вероятностью $1/2 + 2^3(3/16 - 1/2)^2(11/32 - 1/2)^2 = 0{,}519$.

Пусть далее $A_{i,j}$ обозначает ЛСА i-го раунда DES, полученный подстановкой R_{i-1} и K_i вместо x_R и k соответственно и $L_{i-1} \oplus R_i$ вместо y_f в ЛСА функции f, заданный уравнением $(j + 2)$ для $j + 2 = 3, 4, ..., 7$. Непосредственно проверяется, что сумма

$$A = A_{1,5} \oplus A_{3,4} \oplus A_{4,3} \oplus A_{5,1} \oplus A_{7,1} \oplus A_{8,3} \oplus A_{9,4} \oplus A_{11,4} \oplus A_{12,3} \oplus A_{13,1} \oplus A_{15,1}$$

образует ЛСА для 15-раундового DES

$$\begin{aligned} &L_0[7,18,24] \oplus R_0[12,16] \oplus L_{15}[15] \oplus R_{15}[7,18,24,29] = \\ &= K_1[19,23] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus K_7[22] \oplus \\ &\oplus K_8[44] \oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22], \end{aligned} \tag{17}$$

выполняемый с вероятностью $1/2 + 2^{10}(3/16 - 1/2)^4(15/32 - 1/2)^3(21/32 - 1/2)^3(1/4 - -1/2) = 1/2 + 1{,}19 \cdot 2^{-22}$, а сумма $A \oplus A_{16,2}$ есть ЛСА для 16-раундового DES

$$\begin{aligned} &L_0[7,18,24] \oplus R_0[12,16] \oplus L_{16}[7,18,24,27,28,29,30,31] \oplus R_{16}[15] = \\ &= K_1[19,23] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus K_7[22] \oplus K_8[44] \oplus \\ &\oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22] \oplus K_{16}[42,43,45,46], \end{aligned} \tag{18}$$

выполняемый с вероятностью $1/2 + 2(1{,}19 \cdot 2^{-22})(11/32 - 1/2) = 1/2 - 1{,}49 \cdot 2^{-24}$. Кроме того, сумма

$$A_{15,5} \oplus A_{13,4} \oplus A_{12,3} \oplus A_{11,1} \oplus A_{9,1} \oplus A_{8,3} \oplus A_{7,4} \oplus A_{5,4} \oplus A_{4,3} \oplus A_{3,1} \oplus A_{1,1}$$

есть ещё один ЛСА для 15-раундового DES с вероятностью $1/2 + 1{,}19 \cdot 2^{-22}$

$$\begin{aligned} &R_{15}[7,18,24] \oplus L_{15}[12,16] \oplus R_0[15] \oplus L_0[7,18,24,29] = \\ &= K_{15}[19,23] \oplus K_{13}[22] \oplus K_{12}[44] \oplus K_{11}[22] \oplus \\ &\oplus K_9[22] \oplus K_8[44] \oplus K_7[22] \oplus K_5[22] \oplus K_4[44] \oplus K_3[22] \oplus K_1[22], \end{aligned} \tag{19}$$

а сумма

$$A' = A_{16,5} \oplus A_{14,4} \oplus A_{13,3} \oplus A_{12,1} \oplus A_{10,1} \oplus A_{9,3} \oplus A_{8,4} \oplus A_{6,4} \oplus A_{5,3} \oplus A_{4,1} \oplus A_{2,1} \oplus A_{1,2}$$

— ещё один ЛСА 16-раундового DES с вероятностью $1/2 - 1{,}49 \cdot 2^{-24}$

$$\begin{aligned} &R_{16}[7,18,24] \oplus L_{16}[12,16] \oplus R_0[7,18,24,27,28,29,30,31] \oplus L_0[15] = \\ &= K_{16}[19,23] \oplus K_{14}[22] \oplus K_{13}[44] \oplus K_{12}[22] \oplus K_{10}[22] \oplus K_9[44] \oplus \\ &\oplus K_8[22] \oplus K_6[22] \oplus K_5[44] \oplus K_4[22] \oplus K_2[22] \oplus K_1[42,43,45,46]. \end{aligned} \tag{20}$$

Заметим, что последние два ЛСА могут быть получены из (17) и (18) соответственно по следующему правилу, справедливому благодаря симметрии раундов DES: если в ЛСА t-раундового DES произвести взаимную замену L_0, R_0, K_i на R_t, L_t, K_{t+1-i} соответственно для $i = 1, 2, ..., t$, то получится снова ЛСА t-раундового DES с той же вероятностью.

Наконец можно убедиться, что $A - A_{1,5}$ (т. е. A за исключением слагаемого $A_{1,5}$) есть уравнение

$$\begin{aligned} &R_1[7,18,24] \oplus L_{15}[15] \oplus R_{15}[7,18,24,29] = K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus \\ &\oplus K_7[22] \oplus K_8[44] \oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22], \end{aligned} \tag{21}$$

а $A' - A_{16,5} - A_{1,2}$ есть уравнение

$$\begin{aligned} &R_1[15] \oplus L_1[7,18,24,29] \oplus L_{15}[7,18,24] = K_{14}[22] \oplus K_{13}[44] \oplus K_{12}[22] \oplus \\ &\oplus K_{10}[22] \oplus K_9[44] \oplus K_8[22] \oplus K_6[22] \oplus K_5[44] \oplus K_4[22] \oplus K_2[22], \end{aligned} \tag{22}$$

и оба они выполняются с вероятностью $1/2 + 2^9(3/16 - 1/2)^4(15/32 - 1/2)^3(21/32 - -1/2)^3 = 1/2 + 1{,}19 \cdot 2^{-21}$, представляя собой два различных ЛСА для одной и той же функции — 14-раундового DES со 2-го по 15-й раунды. Взаимная замена L_1, R_1, K_i на R_{15}, L_{15}, K_{17-i} соответственно превращает один из них в другой по свойству симметрии раундов DES.

4. Криптоанализ на основе ЛСА

В криптографии этот метод известен как линейный криптоанализ. Применительно к DES его впервые описал японец Mitsuru Matsui [2]. Линейный криптоанализ является атакой с известным открытым текстом, направленной на частичное раскрытие ключа шифра и осуществляемой на основе некоторой системы эффективных линейных статистических аналогов функции шифрования

$$(a^{(i)}, x) \oplus (b^{(i)}, y) = (c^{(i)}, k), i = 1, 2, ..., s, \tag{23}$$

выполнимых с (отличными от $1/2$) вероятностями $p_1, p_2, ..., p_s$ соответственно.

Известные открытые тексты $x^{(j)} \in (\mathbb{Z}_2)^n$ и их криптограммы $y^{(j)} \in (\mathbb{Z}_2)^r$, $j = 1, 2, ..., N$, подставляются в уравнения данной системы, и получается система линейных булевых уравнений для некоторых компонент неизвестного ключа k, а именно:

$$(c^{(i)}, k) = d_{ij}, \quad i = 1, 2, ..., s; \quad j = 1, 2, ..., N, \tag{24}$$

где $d_{ij} = (a^{(i)}, x^{(j)}) \oplus (b^{(i)}, y^{(j)}) \in \mathbb{Z}_2$ для всех $i = 1, 2, ..., s$ и $j = 1, 2, ..., N$. Каждое уравнение в ней вероятностное — по утверждению 3 выполняется с той же вероятностью, что и аналог в (23), из которого оно получено.

Система уравнений (24) относится к классу так называемых случайных систем уравнений с искажённой правой частью [6], ставших в последнее время предметом многочисленных исследований (см., например, «Труды по дискретной математике», издаваемые с 1997 г. совместно Российской академией наук и Академией криптографии РФ как приложение к журналу «Дискретная математика», где можно найти и разные методы решения таких систем). Для решения системы (24) воспользуемся методом максимального правдоподобия (МП).

Пусть $t_i = N - \sum\limits_{j=1}^{N} d_{ij}, i = 1, 2, ..., s$. Это есть количество тех известных пар x/y (открытый текст/криптограмма), для которых левая часть i-го уравнения в (23) обращается в 0. Для каждого $i = 1, 2, ..., s$ определим $d_i \in \mathbb{Z}_2$ по следующим правилам:

1) $d_i = 0$, если $t_i > N/2$ и $p_i > 1/2$ или $t_i \leqslant N/2$ и $p_i < 1/2$;

2) $d_i = 1$, если $t_i \leqslant N/2$ и $p_i > 1/2$ или $t_i > N/2$ и $p_i < 1/2$.

Следуя методу МП, систему (24) заменим детерминированной системой булевых уравнений

$$(c^{(i)}, k) = d_i, \quad i = 1, 2, ..., s, \tag{25}$$

которую можно решить методом Гаусса.

Любое решение любой совместной подсистемы последней системы относительно компонент вектора k, явно входящих в уравнения подсистемы, представляется как результат криптоанализа.

При $m = |k| \leqslant s$ совместная система линейно независимых уравнений (25) имеет 2^{m-s} решений: в них значения некоторых $m - s$ неизвестных выбираются произвольно, а остальные s неизвестных вычисляются по ним однозначно. Это значит, что методом линейного криптоанализа на основе s статистических линейных аналогов шифра в действительности можно определить самое большее s бит ключа. В частности, по одному ЛСА с k_j в правой части находится ровно один ключевой бит — k_j.

Ввиду вероятностного характера уравнений в (23) и (24) результат линейного криптоанализа оказывается также вероятностным: найденные значения компонент ключа являются истинными лишь с некоторой вероятностью. Эта *вероятность успеха* тем

выше, чем выше эффективности использованных статистических линейных аналогов и чем больше открытых текстов занято в атаке. Так, в случае одного ЛСА в системе (23) с вероятностью p и эффективностью $\varepsilon = |p - 1/2| > 0$ для вероятности успеха, близкой к 0,98, требуется около ε^{-2} известных открытых текстов. Например, для нахождения данным методом одного бита ключа в 5-раундовом DES, равного правой части уравнения (16), необходимо иметь $|0{,}519 - 1/2|^{-2} \approx 2800$ открытых текстов. Оба ЛСА (18) и (20) для 16-раундового DES выполняются одновременно с вероятностью $1/2 - 1{,}49 \cdot 2^{-24}$, поэтому линейный криптоанализ на их основе позволяет с большой вероятностью успеха определить сразу два бита ключа 16-раундового DES, используя $(1{,}49 \cdot 2^{-24})^{-2} \approx 2^{47}$ известных открытых текстов.

Основная трудность, с которой сталкивается разработчик метода линейного криптоанализа для конкретного шифра, заключается в построении достаточно большого числа линейных статистических аналогов его функции шифрования с не слишком малой их эффективностью. К сожалению, не много найдётся реальных шифров, для которых такое построение действительно возможно. Более перспективным видится применение в криптоанализе вместо ЛСА нелинейных статистических аналогов функций шифров.

5. Нелинейные статистические аналоги DES

Имея для $(n-1)$-раундового DES линейный статистический аналог $(a, L_0R_0) \oplus \oplus (b', L_{n-1}) \oplus (b'', R_{n-1}) = (c, K)$, выполняемый с некоторой вероятностью p, и равенство $(b', L_{n-1}) = (b', f(L_n, K_n)) \oplus (b', R_n)$, справедливое ввиду (9) при верном раундовом ключе K_n, получаем нелинейный статистический аналог для n-раундового DES

$$(a, L_0R_0) \oplus (b'', L_n) \oplus (b', R_n) \oplus (b', f(L_n, K_n)) = (c, K),$$

выполняемый с той же вероятностью p. Вводя обозначения $b = b''b'$ и $d = b'$, можно переписать последний как

$$(a, L_0R_0) \oplus (b, L_nR_n) \oplus (d, f(L_n, K_n)) = (c, K).$$

Таким способом, например, из ЛСА (17) и (19) для 15-раундового DES ввиду равенств

$$L_{15}[15] = f(L_{16}, K_{16})[15] \oplus R_{16}[15]$$

и

$$L_{15}[7, 18, 24] = f(L_{16}, K_{16})[7, 18, 24] \oplus R_{16}[7, 18, 24]$$

получаются следующие НСА для 16-раундового DES:

$$\begin{gathered} L_0[7, 18{,}24] \oplus R_0[12, 16] \oplus L_{16}[7, 18, 24, 29] \oplus R_{16}[15] \oplus f(L_{16}, K_{16})[15] = \\ = K_1[19, 23] \oplus K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus K_7[22] \oplus \\ \oplus K_8[44] \oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22] \end{gathered} \tag{26}$$

и

$$\begin{gathered} L_0[15] \oplus R_0[7{,}18, 24, 29] \oplus L_{16}[12, 16] \oplus R_{16}[7, 18, 24] \oplus f(L_{16}, K_{16})[7, 18, 24] = \\ = K_{15}[19, 23] \oplus K_{13}[22] \oplus K_{12}[44] \oplus K_{11}[22] \oplus K_9[22] \oplus K_8[44] \oplus \\ \oplus K_7[22] \oplus K_5[22] \oplus K_4[44] \oplus K_3[22] \oplus K_1[22] \end{gathered} \tag{27}$$

соответственно. Каждый из них выполняется с вероятностью $1/2 + 1{,}19 \cdot 2^{-22}$.

Кроме того, имея для $(n-2)$-раундового DES линейный статистический аналог $(a', L_1) \oplus (a'', R_1) \oplus (b', L_{n-1}) \oplus (b'', R_{n-1}) = (c, K)$, выполняемый с некоторой вероятностью p, и равенства $(a'', R_1) = (a'', f(R_0, K_1)) \oplus (a'', L_0)$ и $(b', L_{n-1}) = (b', f(L_n, K_n)) \oplus \oplus (b', R_n)$, справедливые ввиду (9) при верных раундовых ключахK_1 и K_n, получаем нелинейный статистический аналог для n-раундового DES

$$(a'', L_0) \oplus (a', R_0) \oplus (a'', f(R_0, K_1)) \oplus (b'', L_n) \oplus (b', R_n) \oplus (b', f(L_n, K_n)) = (c, K),$$

выполняемый с той же вероятностью p.

Так, из уравнения (21) и равенств

$$R_1[7, 18, 24] = f(R_0, K_1)[7, 18, 24] \oplus L_0[7, 18, 24], \quad L_{15}[15] = f(L_{16}, K_{16})[15] \oplus R_{16}[15]$$

и из уравнения (22) и равенств

$$R_1[15] = f(R_0, K_1)[15] \oplus L_0[15], \quad L_{15}[7, 18, 24] = f(L_{16}, K_{16})[7, 18, 24] \oplus R_{16}[7, 18, 24]$$

получаются следующие НСА для 16-раундового DES, выполняемые с вероятностью $1/2 + 1{,}19 \cdot 2^{-21}$:

$$\begin{aligned} L_0[7, 18, 24] \oplus f(R_0, K_1)[7, 18, 24] \oplus L_{16}[7, 18, 24, 29] \oplus R_{16}[15] \oplus f(L_{16}, K_{16})[15] = \\ = K_3[22] \oplus K_4[44] \oplus K_5[22] \oplus K_7[22] \oplus K_8[44] \oplus \\ \oplus K_9[22] \oplus K_{11}[22] \oplus K_{12}[44] \oplus K_{13}[22] \oplus K_{15}[22] \end{aligned} \tag{28}$$

и

$$\begin{aligned} L_0[15] \oplus f(R_0, K_1)[15] \oplus R_0[7, 18, 24, 29] \oplus R_{16}[7, 18, 24] \oplus f(L_{16}, K_{16})[7, 18, 24] = \\ = K_{14}[22] \oplus K_{13}[44] \oplus K_{12}[22] \oplus K_{10}[22] \oplus K_9[44] \oplus \\ \oplus K_8[22] \oplus K_6[22] \oplus K_5[44] \oplus K_4[22] \oplus K_2[22] \end{aligned} \tag{29}$$

соответственно.

По определению раундовой функции f функция $f(X, K)[15]$ реализуется на выходе S-блока S_1, а функция $f(X, K)[7, 18, 24]$ — на выходах S-блока S_5. Тем самым каждая из этих двух функций существенно зависит только от шести бит раундового ключа K: первая — от $K[42], K[43], ..., K[47]$, вторая — от $K[18], K[19], ..., K[23]$. Это значит, что в приведенных выше уравнениях (26), (27) и (28), (29) нелинейные слагаемые зависят на самом деле соответственно от 6 и от 12 неизвестных ключевых бит, а именно: в (26) — от $K_{16}[42], K_{16}[43], ..., K_{16}[47]$; в (27) — от $K_{16}[18], K_{16}[19], ..., K_{16}[23]$; в (28) — от $K_1[18], K_1[19], ..., K_1[23], K_{16}[42], K_{16}[43], ..., K_{16}[47]$; в (29) — от $K_1[42], K_1[43], ..., K_1[47], K_{16}[18], K_{16}[19], ..., K_{16}[23]$.

6. 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, однако в отличие от линейного он 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: 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 нелинейный статистический аналог $\varphi(x, y, k) = 0$ функции шифрования $F(x, k)$, выполняемый с некоторой вероятностью p, и $0 < p < 1$.

6.1. 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 её 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 сумма значений переменных в 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 $d_{ij} = \varphi'(x^{(i)}, y^{(i)}, k'^{(j)})$, после чего для каждого $k'^{(j)}$ подсчитывается количество $t_j = N - \sum\limits_{i=1}^{N} d_{ij}$ всех таких пар $(x^{(i)}, y^{(i)})$, для которых $d_{ij} = 0$, и определяются m и l из условий: $t_m = \max t_j$ и $t_l = \min t_j$ по всем j от 1 до N.

2. Если $|t_m - N/2| > |t_l - N/2|$, то полагаем $k' = k'^{(m)}$ и, кроме того, $d = 0$ в случае $p > 1/2$ и $d = 1$ в случае $p < 1/2$. Если же $|t_m - N/2| < |t_l - N/2|$, то полагаем $k' = k'^{(l)}$ и, кроме того, $d = 1$ для $p > 1/2$ и $d = 0$ для $p < 1/2$.

(Иначе говоря, 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 ещё один бит информации о ключе k.

К сожалению, мы не знаем, зависят ли статистически от переменных в $x'k'$ функции, ассоциированные с приведёнными выше НСА (26) — (29) для DES, и, следовательно, не знаем, являются ли последние разделимыми по переменным. В соответствии с нашей теорией это значит, что мы не вправе использовать эти НСА в данном алгоритме, поскольку нет гарантии того, что вероятность выполнения уравнения, по-

лученного подстановкой символов открытого и соответствующего шифрованного текстов в любой из них, не будет сильно отличаться от его вероятности p и не совпадёт с $1/2$. Сам M. Matsui признаёт в [3], что эта вероятность «is expected to be closer to $1/2$ (not necessarily $1/2$)», и тем не менее с верой в не только русское «наука полагает, а Бог располагает» применяет алгоритм с каждым из linear approximate equations (26) — (29) в криптоанализе DES.

Согласно [2], количество N известных открытых текстов, при котором вероятность успеха данного алгоритма в применении к DES близка к 0,97, оценивается величиной $8\varepsilon^{-2}$. Так, применив его с $N = 8(1{,}19 \cdot 2^{-22})^{-2} \approx 2^{47}$ известными открытыми текстами дважды: сначала — к НСА (26), затем — к НСА (27), M. Matsui находит 14 бит ключа DES, а именно: $K_{16}[42], K_{16}[43], ..., K_{16}[47]$, один бит из правой части в (26), $K_{16}[18], K_{16}[19], ..., K_{16}[23]$ и один бит из правой части в (27). Аналогичным образом с использованием $N = 8(1{,}19 \cdot 2^{-21})^{-2} \approx 2^{45}$ известных открытых текстов по (28) и (29) находятся 26 бит ключа DES, а именно: $K_1[18], K_1[19], ..., K_1[23], K_{16}[42], K_{16}[43], ..., K_{16}[47]$, один бит из правой части в (28), $K_1[42], K_1[43], ..., K_1[47], K_{16}[18], K_{16}[19], ..., K_{16}[23]$ и один бит из правой части в (29).

6.2. 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:

$$\varphi_i(k) = 0, \quad i = 1, 2, ..., N, \tag{30}$$

где $\varphi_i(k) = \varphi(x^{(i)}, y^{(i)}, k)$ для всех $i \in \{1, 2, ..., N\}$. Каждое уравнение в системе (30) вероятностное и выполняется с той же вероятностью p, что и НСА, из которого оно получено.

Следуя [7], назовём подмножество переменных в k *линеаризационным*, если при фиксации любых их значений каждое уравнение в системе (30) превращается в линейное (линеаризуется). Зафиксируем некоторое (лучше — наименьшей мощности, или кратчайшее) линеаризационное множество L переменных в системе (30). Для каждого набора $L^{(j)}$ значений переменных в L возьмём 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

$$(c_i^{(j)}, k') = d_i^{(j)}, \quad i = 1, 2, ..., N; j = 1, 2, ..., s, \tag{31}$$

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 (31) для $i = 1, 2, ..., N$. Каждая подсистема E_j решается подобно системе (24), а именно: полагаем $t_j = N - \sum\limits_{i=1}^{N} d_i^{(j)}$, определяем $d^{(j)}$ по следующим правилам:

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

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

и записываем детерминированную систему уравнений

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

Если эта система совместна, то её 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. Легко видеть, что каждый из приведённых выше НСА (26), (27), (28) и (29) для 16-раундового DES этим свойством обладает, ибо множества $L_1 = \{K_{16}[42], K_{16}[43], ..., K_{16}[47]\}$, $L_2 = \{K_{16}[18], K_{16}[19], ..., K_{16}[23]\}$, $L_3 = \{K_1[18], K_1[19], ..., K_1[23], K_{16}[42], K_{16}[43], ..., K_{16}[47]\}$ 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 (30) для этих НСА соответственно. Таким образом, применив данный алгоритм с НСА (26) или с НСА (27), можно получить 18 бит ключа DES: 6 бит в L_1 и 12 бит из правой части в (26) или 6 бит в L_2 и 12 бит из правой части в (27) соответственно. Применив же его с НСА (28) или с НСА (29), можно получить 22 бита ключа DES, а именно: 12 бит в L_3 и 10 бит из правой части в (28) или 12 бит в L_4 и 10 бит из правой части в (29). Если же применить алгоритм сначала, скажем, с НСА (28), а затем с НСА (29), то можно получить 44 бита раундовых ключей DES, или, с учётом расписания ключей, 34 бита исходного ключа, в то время как M. Matsui находит от тех же самых двух НСА только 26 из этих бит.

Заметим, что если функция φ является сильно t-аффинной [8], то система (30) имеет линеаризационное множество мощности t, поэтому для противостояния этой атаке необходимо, чтобы функция шифрования не допускала статистического аналога с функцией, имеющей малый уровень сильной аффинности.

Атака с выбором открытого текста

Мощность линеаризационного множества переменных в системе (30) 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 (30), 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 её ускорение возможно на пути выбора более подходящего НСА. Впрочем, последнее замечание относится и к атаке с известным открытым текстом.

ЛИТЕРАТУРА

1. *Агибалов Г. П.* Методы решения систем уравнений над конечным полем // Вестник Томского госуниверситета. Приложение. 2006. № 17. С. 4–9.
2. *Matsui M.* Linear Cryptanalysis Method for DES Cipher // LNCS. 1993. V. 765. P. 386–397.
3. *Matsui M.* The First Experimental Cryptanalysis of the Data Encryption Standard // LNCS. 1994. V. 839. P. 1–11.
4. *Логачев О. А., Сальников А. А., Ященко В. В.* Булевы функции в теории кодирования и криптографии. М.: МЦНМО, 2004.

5. *Агибалов Г. П.* Элементы теории дифференциального криптоанализа итеративных блочных шифров с аддитивным раундовым ключом // Прикладная дискретная математика. 2008. № 1. С. 34–43.
6. *Балакин Г. В.* Введение в теорию случайных систем уравнений // Труды по дискретной математике. Т. 1. М.: ТВП, 1997. С. 1–18.
7. *Агибалов Г. П.* Логические уравнения в криптоанализе генераторов ключевого потока // Вестник Томского госуниверситета. Приложение. 2003. № 6. С. 31–41.
8. *Буряков М. Л., Логачев О. А.* Об уровне аффинности булевых функций // Дискретная математика. 2005. Т. 17. Вып. 4. С. 98–107.