Ростовцев А. Г., Маховенко Е. Б.

ДВА ПОДХОДА К АНАЛИЗУ БЛОЧНЫХ ШИФРОВ

Предложены два эвристических подхода к вскрытию ключа итерированного блочного шифра (обращения хэш-функции). Первый подход основан на вложении кольца G_n многочленов Жегалкина в квазикольцо рациональных чисел, второй — на вложении G_n в (квази)кольцо целых 2-адических чисел. Для вскрытия ключа необходимо составить продолженную целевую функцию и найти ее максимумы методом наискорейшего спуска. Значения переменных, найденные в ходе поиска максимума, дают предположительно правильные разряды ключа.

Подходы подтверждены экспериментально. Для ГОСТ 28147–89 найден обширный класс потенциально слабых ключей, которые, возможно, допускают вскрытие с низкой сложностью. При этом требуется всего четыре блока подобранных открытых текстов.

1. Итерированные криптоалгоритмы и методы их анализа

Криптография позволяет создавать методы защиты информации, обеспечивающие безопасность в условиях невырожденной модели возможностей нарушителя (если чтение области памяти не является единственным способом узнать защищаемую информацию) [6]. При разработке криптографического алгоритма наиболее трудоемкой задачей является обоснование его безопасности — оценка сложности вскрытия ключа.

В настоящее время для обеспечения конфиденциальности обычно используются симметричные итерированные блочные шифры. Безопасность таких шифров основана на задаче вскрытия ключа по известным или подобранным открытым и зашифрованным текстам. Сложность этой задачи обусловлена тем, что трудно определить, насколько тестируемый ключ близок к истинному, так как небольшое изменение ключа или открытого текста вызывает значительное изменение шифрограммы. Иначе говоря, трудно задать вычислимую метрику, показывающую “расстояние” между тестируемым и истинным ключом.

Для обеспечения высокой стойкости число циклов шифрования принято выбирать достаточно большим. Однако до настоящего времени никто не доказал, что увеличение числа циклов шифрования повышает стойкость шифра или хотя бы не снижает ее; это предположение является своего рода криптографическим “фольклором” (и не всегда справедливо).

Универсальные методы криптоанализа можно разделить на статистические и алгебраические. Статистические методы используют метрику, характерную для текстов “в среднем”, при этом обычно требуется большой (зачастую недоступный) объем статистики открытых и соответствующих зашифрованных текстов. Сюда можно отнести дифференциальный [8] и линейный [11] методы, метод списка ключей [10] и метод сдвига [9], а также ряд других методов, опубликованных в зарубежной печати. Чтобы

организовать противодействие таким методам достаточно периодически менять ключи, не позволяя нарушителю набрать требуемый объем статистики.

Алгебраические методы часто основаны на использовании принципа Пойа "обобщение — редукция" [2] и не требуют большого объема статистики, поэтому организационно противостоять им почти невозможно. Эти методы позволяют не только вскрывать ключ шифра, но и находить аргумент хэш-функции. Сюда относятся метод Андельмана — Ридса [7] и решеточный метод [4]. Сложность первого метода растет по экспоненте от числа циклов шифрования, что обусловлено необходимостью вычисления производной сложной функции. Модификация этого метода предложена в работе [3]. К числу алгебраических можно отнести и метод "giant step — baby step" для вскрытия ключа симметричного шифра (реализуемый на квантовом компьютере) и этим снизить стойкость в 2 раза по порядку величины [1]. Однако возможность создания практического квантового компьютера неочевидна.

При анализе алгебраическими методами нужно задать вычислимую целевую функцию, принимающую значение 1 на единственном наборе переменных, продолжить ее до некоторого упорядоченного множества и найти ее экстремум. В качестве целевой функции можно использовать поразрядную конъюнкцию зашифрованных текстов, вычисленных для известного открытого текста и тестируемого ключа, и истинного значения шифрограммы. На практике целевую функцию нельзя записать в виде обычной булевой формулы, так как эта формула очень сложна, но можно вычислить с полиномиальной сложностью для любого набора переменных.

2. Кольцо многочленов Жегалкина и его продолжения

Любую булеву функцию можно однозначно задать в виде многочлена Жегалкина от n переменных $x_1, \ldots, x_n$. Многочлены Жегалкина образуют конечное ассоциативное коммутативное факториальное кольцо $\boldsymbol{G_n}$ характеристики 2, в котором каждый элемент, отличный от константы, является делителем нуля. Кольцо $\boldsymbol{G_n}$ содержит в точности n неразложимых многочленов, которые принимают единственное нулевое значение на множестве из 2^n наборов переменных [6], например, $1 \oplus x_1 \ldots x_n$, где $\oplus$ — сложение по модулю 2.

Имеет место изоморфизм колец [6]:

$$\boldsymbol{G_n} \cong \mathbf{F}_2[x_1, \ldots, x_n]/(x_1^2 \oplus x_1, \ldots, x_n^2 \oplus x_n).$$

Замена всех переменных значениями из $\mathbf{F}_2$ задает гомоморфизм колец $\boldsymbol{G_n} \to \mathbf{F}_2$, который можно рассматривать как эндоморфизм кольца $\boldsymbol{G_n}$.

Для анализа шифров интересны такие продолжения кольца $\boldsymbol{G_n}$ на упорядоченные множества, которые позволяют определить, насколько тес-

тируемый разряд ключа близок к истинному значению. Предлагаются два подхода к анализу блочных шифров и хэш-функций.

Первый подход использует продолжение $\boldsymbol{G_n}$ на кольцо многочленов над полем рациональных чисел $\mathbf{Q}$ (вещественные числа, не являющиеся рациональными, непредставимы в ЭВМ). Более строго можно говорить не о продолжении кольца $\boldsymbol{G_n}$, а о продолжении эндоморфизмов этого кольца.

Наиболее интересным представляется продолжение

$$a \oplus b \to |a - b|,\ ab \pmod 2 \to ab, \qquad (1)$$

где $\oplus$ — сложение по модулю 2. Продолжение операции $\oplus$ определяет коммутативную операцию "сложения", задающую на рациональных числах структуру, похожую на коммутативную группу (без ассоциативности и однозначной разрешимости уравнений). Продолженная структура удовлетворяет многим аксиомам кольца и имеет характеристику 2. Назовем эту структуру *квазикольцом*.

Если переменные принимают значения из подмножества рациональных чисел от 0 до 1, то и продолженный многочлен принимает значения из этого множества.

Второй подход предполагает продолжение эндоморфизмов $\boldsymbol{G_n} \to \mathbf{F}_2$ до эндоморфизмов $\mathbf{Z}_2[x_1, \ldots, x_n] \to \mathbf{Z}_2$, где $\mathbf{Z}_2$ — кольцо целых 2-адических чисел. Значение нормирования целого числа $c2^k$, где c — нечетное число, равно $-k$. Таким образом, значение нормирования является целым неположительным числом. Удвоение целого числа ведет к уменьшению значения его нормирования на 1. Такое продолжение арифметических операций задает гомоморфное вложение $\boldsymbol{G_n}$ в $\mathbf{Z}_2[x_1, \ldots, x_n]$. Наряду с указанным возможно и продолжение $a \oplus b \to |a - b|$. При этом продолженная структура 2-адических чисел будет являться не кольцом, а квазикольцом, но значение нормирования удвоенного числа уменьшается до $-\infty$.

По аналогии с приближением вещественных чисел рациональными, когда отбрасываются младшие разряды, целое 2-адическое число можно приближенно представлять элементом кольца $\mathbf{Z}/2^m\mathbf{Z}$. В этом случае отбрасываются старшие двоичные разряды целого числа, минимальное значение нормирования равно $-m$ и продолжение имеет вид

$$a \oplus b \to a + b \pmod{2^m},\ ab \pmod 2 \to ab \pmod{2^m}. \qquad (2)$$

Иногда сложение лучше продолжать так:

$$a \oplus b \to |a - b| \pmod{2^m}. \qquad (2a)$$

Абсолютная величина разности обеспечивает коммутативность продолженного сложения, но для целей анализа можно использовать просто разность.

3. Метод анализа

В основе криптоанализа лежит процедура нахождения максимума целевой функции H, которая в случае продолжения (1) принимает значение 1 для истинного значения ключа. В случае продолжения (2), (2а) в качестве значения целевой функции используется значение ее дискретного нормирования, равное нулю для истинного значения ключа. Поскольку аппарат дифференцирования для нахождения максимума в данном случае неэффективен, использовался метод наискорейшего спуска, при этом множество возможных значений разрядов ключа выбиралось по аналогии с работой [3]. В случае рационального продолжения — это {0; 0,5; 1}, а в случае 2-адического продолжения — {0; 1; 2}. Здесь числа 0,5 и 2 являются промежуточными значениями между нулем и единицей соответственно для первого и второго подхода.

Пусть длина ключа равна n бит. Ключ определен однозначно, если известны в среднем не менее $1{,}36n$ бит открытого и соответствующего зашифрованного текста [5]. В качестве целевой функции шифра с открытым текстом **x** и шифрограммой **y** целесообразно использовать конъюнкцию

$$H = \bigwedge_{i=1}^{O(n)} (u_i \oplus v_i \oplus 1), \tag{3}$$

где u_i — разряды промежуточного текста, полученные при зашифровании открытого текста на половине циклов шифрования, v_i — разряды промежуточного текста, полученные при расшифровании шифрограммы на оставшейся половине циклов шифрования.

Целевая функция (3), продолженная на поле **Q** согласно (1), имеет вид

$$H = \prod_{i=1}^{O(n)} \left(1 - |u_i - v_i|\right) \tag{4}$$

или

$$H = \prod_{i=1}^{O(n)} \left|u_i + v_i - 1\right|. \tag{5}$$

Если переменные принимают значения из интервала (0, 1), то и целевые функции (4), (5) принимают значения из этого же интервала.

Целевая функция (3) в случае 2-адического продолжения (2), (2а) имеет вид

$$H = \prod_{i=1}^{O(n)} \left(1 + u_i + v_i\right) (\mathrm{mod}\, 2^m) \tag{6}$$

или

$$H = \prod_{i=1}^{O(n)} \left(1 + |u_i - v_i|\right) (\mathrm{mod}\, 2^m). \tag{7}$$

Очевидно, что если все разряды ключа определены правильно, то значение продолженной целевой функции (ПЦФ) будет максимально. Для рационального продолжения (4) или (5) это значение равно 1, а для 2-адического — 0, что соответствует произвольному нечетному значению H в (6) или (7).

Продолжения многочленов Жегалкина нарушают ряд алгебраических свойств исходного кольца многочленов, поэтому у ПЦФ появляются локальные экстремумы, не соответствующие истинному значению ключа. Опытным путем установлено, что двоичные значения разрядов шифра, выявленные в ходе поиска локального максимума ПЦФ, обычно чаще оказываются правильными, чем неправильными.

Поиск локального максимума ПЦФ является не целью, а средством выявления предположительно истинных разрядов ключа. Наряду с поиском максимума ПЦФ можно использовать другие критерии. Способ продолжения эндоморфизмов кольца $\boldsymbol{G_n}$, критерии выбора предположительно правильного разряда ключа, а также способ представления операций в виде многочленов Жегалкина выбираются в зависимости от вида шифра.

Метод анализа предполагает три основных этапа. На этапе предвычислений изучаются статистические свойства ПЦФ. Для случаев, когда ключ известен, выбираются способы продолжения эндоморфизмов $\boldsymbol{G_n}$ и задания целевой функции. Для частоты p совпадения найденного разряда ключа с истинным значением вычисляется преобладание $\varepsilon = p - 0{,}5$. На втором этапе по известным открытым и зашифрованным текстам оценивается ключ как совокупность наиболее часто встречающихся разрядов. На третьем этапе выполняется опробование ключей, близких к найденной оценке. Все этапы допускают распараллеливание.

Сложность вскрытия ключа существенно зависит от выполнения следующего предположения.

Предположение 1. Каждый разряд ключа вскрывается независимо от остальных с положительным преобладанием ε.

Если предположение 1 верно, то сложность вскрытия одного разряда ключа составляет примерно ε^{-2} итераций. Если ε^{-2} оценивается многочленом от n, то третий этап оказывается лишним, а сложность вскрытия ключа становится полиномиальной.

Если предположение 1 неверно, то предложенный метод благодаря этапу 3 может иметь сверхполиномиальную сложность, но остается менее сложным, чем перебор.

Обозначим через $p_{i/j}$ частоту того, что для данного бита ключа найдена оценка i, тогда как в действительности этот бит имеет значение j. Если предположение 1 не выполняется, то истинное значение ключа следует искать перебором вблизи найденной оценки. Обозначим через n_0, n_1 соот-

ветственно число нулей и единиц в оценке ключа, $n_{00} = p_{0/0}n_0$, $n_{01} = p_{0/1}n_0$, $n_{10} = p_{1/0}n_1$, $n_{11} = p_{1/1}n_1$. Объем перебора можно оценить числом $\binom{n_0}{n_{01}}\binom{n_1}{n_{10}}$.

Эта оценка предполагает, что все нулевые разряды ключа могут быть ошибочными с одинаковой вероятностью и все единичные разряды тоже могут быть ошибочными с одинаковой вероятностью. На практике объем перебора можно снизить, если по результатам первого этапа разбить множество разрядов на классы по частоте ошибок и вести перебор внутри соответствующих классов.

4. Результаты эксперимента

Предложенные подходы были проверены экспериментально на нескольких шифрах и хэш-функциях. Первый вариант исследуемого шифра имел длину блока и ключа по 64 бита и задавался уравнением $\mathbf{y} = F^{2k}(\mathbf{x})$, где F — оператор шифрования на каждом из $2k$ циклов, содержащий сложение по модулю 2 текста с ключом, перестановку битов в блоке, задаваемую уравнением $x_i \to x_{23i(\mathrm{mod}\ 64)}$, экстремальную 4-битовую подстановку

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 0 & 13 & 11 & 8 & 3 & 6 & 4 & 1 & 15 & 2 & 5 & 14 & 10 & 12 & 9 & 7 \end{pmatrix},$$

обладающую наилучшими теоретически возможными характеристиками (нелинейность равна 4, диффузия равна 1 [6]), и циклический сдвиг на 25 бит. Абсолютная величина преобладания любого линейного равенства $\sum_i x_i + \sum_j y_j = 0$ (разность между 0,5 и вероятностью выполнения этого равенства над полем $\mathbf{F}_2$) для подстановки не превышает $^1/_4$. Вероятность наиболее вероятного дифференциала подстановки равна $^1/_4$. Используемые операторы шифрования обеспечивают стойкость к дифференциальному и линейному методам близкую к максимально возможной для подстановочно-перестановочных шифров.

При $k = 4$ стойкость этого шифра к линейному методу анализа составляет примерно 2^{38}, столько же требуется открытых и зашифрованных текстов. Наиболее вероятные дифференциалы подстановки имеют вероятность не более $^1/_4$. С учетом того, что число дифференциалов с вероятностью $^1/_4$ больше, чем число линейных сумм с абсолютной величиной преобладания $^1/_4$, для вскрытия ключа дифференциальным методом требуется лишь 2^{35} подобранных открытых текстов (с заданной входной разностью).

При $k = 8$ стойкость шифра к линейному и дифференциальному методам анализа превышает переборную и требует невозможного объема статистики.

Промежуточные тексты имели вид: $\mathbf{u} = F^k(\mathbf{x})$ и $\mathbf{v} = F^{-k}(\mathbf{y})$ для одного известного открытого и соответствующего зашифрованного блока.

Для снижения погрешности аппроксимации (1) и (2а), вызванной неассоциативностью, использовалось нетрадиционное представление булевых функций многочленами Жегалкина.

В ходе выполнения первого этапа для рационального и 2-адического продолжения автоморфизмов кольца $\boldsymbol{G_n}$ установлено, что преобладание достаточно велико для того, чтобы сложность вскрытия ключа была значительно ниже переборной.

Аналогичные результаты получились при вскрытии ключа шифра с попарно различными циклами шифрования и при обращении хэш-функции $\mathbf{y} = F^{2k}(\mathbf{x}) \oplus \mathbf{x}$. Во втором случае в целевой функции использовались промежуточные тексты $\mathbf{u} = F^{k}(\mathbf{x})$, $\mathbf{v} = F^{-k}(\mathbf{x} \oplus \mathbf{y})$. Установлено, что преобладания для хэш-функции и шифра с тем же числом циклов шифрования примерно совпадают. Это обстоятельство опровергает мнение о том, что найти коллизию хэш-функции легче, чем обратить ее.

Эксперимент проводился также с отечественным 32-цикловым стандартом шифрования ГОСТ 28147–89. На сегодняшний день не опубликованы методы, снижающие его стойкость по сравнению с перебором. Существуют классы ключей, при которых шифр становится степенным: это ключи, определяемые равенствами $K_0 = K_7$, $K_1 = K_6$, $K_2 = K_5$, $K_3 = K_4$. Для степенных шифров возможен анализ методом сдвига на один период чередования ключей [9]. Однако эти равенства эффективно проверяемы только при достаточно малом периоде повторения ключей (1 или 2) и при наличии примерно 2^{32} открытых и соответствующих зашифрованных текстов. Это обстоятельство затрудняет использование свойства периодичности на практике.

В эксперименте выполнялся первый этап метода для продолжения (1) эндоморфизмов $\boldsymbol{G_n}$ и целевой функции (4), при этом использовалась экстремальная подстановка из предыдущего примера.

Перемешивающие и рассеивающие свойства шифра в значительной степени определяются переносами при сложении по модулю 2^{32}. Если ключ и текст являются разреженными, то влияние переносов на начальных циклах шифрования ослабляется. В целевой функции использовались четыре блока открытого текста и соответствующие шифрограммы, суммарная длина открытых текстов равна длине ключа. Для ослабления влияния переносов один из блоков был нулевым, а остальные содержали по одной единице. Шифрограммы вычислялись для режима простой замены.

Установлено, что с увеличением разреженности слов ключа оценка преобладания возрастает. Полученные большие оценки преобладания позволяют предположить, что ГОСТ 28147–89 обладает большим множеством потенциально слабых ключей, для которых число единиц в каждом слове не превышает по крайней мере 8, а подстановка сохраняет 0 неподвижным (верхняя граница опасной разреженности не определялась). По-

видимому, такие ключи могут быть вскрыты значительно быстрее, чем перебором. Число таких ключей превышает 10^{57}. Уточнение границ множества потенциально слабых ключей, как и уточнение оценки стойкости стандарта, требует дополнительных исследований.

В результате эксперимента установлено следующее.

1. Сложность предложенного метода практически не зависит от того, одинаковы или попарно различны все циклы шифрования.
2. Значение преобладания, определяющее сложность метода, не является монотонной функцией числа циклов шифрования: увеличение числа циклов может снижать стойкость.
3. Значения преобладаний в случае рационального и 2-адического продолжения для ПЦФ, вычисляемых согласно (4) и (5), а также (6) и (7), различаются несущественно.

Проведенные эксперименты позволяют предположить, что предложенный метод анализа является весьма эффективным, по крайней мере, для подстановочно-перестановочных шифров и хэш-функций, зачастую имеет значительно меньшую сложность, чем известные, и требует несравнимо меньшего объема известных или подобранных текстов. К числу недостатков можно отнести эвристический характер оценок стойкости.

Литература

1. Манин Ю. И. Классическое вычисление, квантовое вычисление и алгоритм факторизации Шора // Квантовый компьютер и квантовые вычисления, т. II. — Ижевск: редакция журнала “Регулярная и хаотическая механика”, 1999. С. 248–286.
2. Пойа Д. Математическое открытие. — М.: Наука, 1976.
3. Ростовцев А. Г. Метод обращения итерированной хэш-функции // Тезисы докладов конференции “Методы и технические средства обеспечения безопасности информации”. — СПб: Изд-во СПбГТУ, 2001. С. 114–117.
4. Ростовцев А. Г. Решеточный криптоанализ // Безопасность информационных технологий, 1997. Вып. 2. С. 53–55.
5. Ростовцев А. Г., Маховенко Е. Б. Введение в криптографию с открытым ключом. — СПб.: Мир и Семья, 2001.
6. Ростовцев А. Г., Маховенко Е. Б. Введение в теорию итерированных шифров. — СПб.: Мир и Семья, 2002.
7. Andelman D., Reeds J. On the cryptanalysis of rotor machines and substitution-permutation networks // IEEE transactions on information theory, v. IT-28, 1982, pp. 578–584.

8. Biham E., Shamir A. Differential cryptanalysis of DES-like cryptosystems // Advances in Cryptology — CRYPTO ′90. LNCS, v. 537, Springer-Verlag, 1991, pp. 2–21.
9. Biryukov A., Wagner D. Slide attacks // Fast software encryption FSE'99, LNCS, v. 1636, 1999, pp. 245–259.
10. Kelsey J., Schneier B., Wagner D. Key-schedule cryptanalysis of IDEA, G-DES, GOST, SAFER, and Triple-DES // Advances in Cryptology — CRYPTO ′96, LNCS, v. 1109, Springer-Verlag, 1996, pp. 237–251.
11. Matsui M. Linear cryptanalysis method for DES cipher // Advances in Cryptology — EUROCRYPT ′93, LNCS, v. 765, 1994, pp. 386–397.