А. Г. Ростовцев

РЕШЕТОЧНЫЙ КРИПТОАНАЛИЗ

Предлагается метод анализа симметричных шифров, полученных путем композиции наборов несложных булевых функций. Метод основан на использовании решеточного продолжения булевых функций с {0, 1} на множество рациональных чисел, лежащих в диапазоне от 0 до 1, и позволяет распознавать, каким именно разрядом ключа определяется значение решеточно продолженной функции.

1. Введение

Большинство современных шифров основано на использовании композиции наборов булевых функций, каждая из которых имеет несложное описание и может быть легко вычислена. Примерами являются DES, ГОСТ 28147-89, FEAL, «Кобра» и т. п. Сложность раскрытия ключа таких шифров обусловлена тем, что небольшое изменение ключа или текста приводит к большому изменению функции, причем неясно, каким разрядом переменной (ключа) определяется значение функции. Существующие методы криптоанализа, в том числе линейный [1] и дифференциальный [2], требуют большого объема известных открытых текстов. В данной работе предложен принцип исследования булевых уравнений, основанный на распознавании разрядов ключа и не требующий большого объема текстов.

Рассмотрим шифратор с композициями наборов булевых функций, без эквивалентных ключей. Предположим, что нарушитель знает криптоалгоритм и некоторое количество разрядов открытых и соответствующих зашифрованных текстов. Пусть разряды открытого текста описываются уравнениями

$$x_i = F_i(y_1, \ldots, y_n, z_1, \ldots, z_N). \qquad (1)$$

Здесь y_i — разряды шифрограммы, z_j — разряды ключа. Булевы функции F_i не имеют аналитической записи (она очень сложна), но значения их могут быть вычислены для произвольного набора аргументов. Если открытые и зашифрованные тексты известны в достаточном количестве, равном $O(N)$, то система уравнений (1) имеет единственное решение, которое и является ключом. При этом достаточно знать только отдельные разряды некоторых блоков открытого текста. Запишем для этого случая систему уравнений (1) в виде

$$f_i(z_1, \ldots, z_N) = a_i \qquad (2)$$

Поскольку система (2) имеет единственное решение, оно может быть записано в виде булевой формулы, содержащей единственную конъюнкцию:

$$\tilde{z}_1\tilde{z}_2\ldots\tilde{z}_N = 1 \qquad (3)$$

«Безопасность информационных технологий»,
М., 1997. Вып. 2. С. 53–55

Символ $\tilde{z}$ означает вхождение переменной с инверсией или без инверсии.

С другой стороны, (2) может быть записана в виде следующей конъюнкции булевых формул:

$$\wedge \tilde{F}_i = 1. \qquad (4)$$

Если $x_i = 1$, то F_i входит без инверсии, в противном случае F_i входит с инверсией. Поскольку левые части выражений (3) и (4) равны 1 на одном и том же наборе аргументов, они представляют одну и ту же булеву функцию, которую назовем *целевой*, при этом (4) дает способ вычисления целевой функции для произвольного ключа.

2. Решеточно продолженные булевы функции

Булевы формулы в функциональном базисе И, ИЛИ, НЕ можно рассматривать не только как элементы булевой алгебры, но и как решетки. При этом множество значений аргументов и функций {true, false} интерпретируется как true = 1, false = 0 с упорядочиванием 1 > 0. Соответственно элементарные функции определяются численно на {0, 1} так:

$$a \vee b = \max(a,b);\ a \wedge b = \min(a,b);\ \bar{a} = 1 - a. \qquad (5)$$

Здесь одной и той же буквой обозначается имя аргумента и его численное значение, но это не вносит путаницу.

Выражения (5) позволяют формально определить произвольную булеву формулу на множестве $\boldsymbol{A}$ чисел, лежащих в интервале от 0 до 1. Множество $\boldsymbol{A}$ может состоять из конечного или счетного числа элементов. При этом формальное представление булевой функции формулой сохраняется, но расширяется ее область определения и, соответственно, область значений. Такие функции будем называть *решеточно продолженными* (РПБФ). Будем называть для РПБФ по аналогии с булевыми формулами $a \vee b$ дизъюнкцией, $a \wedge b$ конъюнкцией. Соответственно определяется и дизъюнктивная нормальная форма.

Геометрически введенные понятия можно интерпретировать так. Булева функция определена на вершинах N-мерного единичного куба, а РПБФ определена на всех рациональных точках этого куба. Нет необходимости расширять множество $\boldsymbol{A}$ до вещественных чисел, так как вещественные нерациональные числа не представимы в ЭВМ.

РПБФ обладает следующими свойствами:

1. Булева функция и ее решеточное продолжение на вершинах единичного куба совпадают.
2. Решеточное продолжение операций И, ИЛИ сохраняет идемпотентность: $a \vee a = a$, $a \wedge a = a$. Поэтому РПБФ принимает значения одной из переменных или ее инверсии.

3. Решеточные продолжения операций И, ИЛИ являются коммутативными, ассоциативными, дистрибутивными. Для них выполняются законы поглощения: $a \vee b = b \vee a$, $a \wedge b = b \wedge a$, $(a \vee b) \vee c = a \vee (b \vee c)$, $(a \wedge b) \wedge c = a \wedge (b \wedge c)$, $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$, $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$, $a \vee (a \wedge b) = a$, $a \wedge (a \vee b) = a$.
4. Операция инверсии не является решеточной операцией дополнения на внутренних точках куба.
5. Для решеточного продолжения операции сложения по модулю 2 слагаемое, ближайшее к числу 0.5, поглощает другие слагаемые суммы. Число 0.5 является аннулятором: $0.5 \oplus z = 0.5$.
6. Отображение $z \to 0$ для $0 \le z < 0.5$ и $z \to 1$ для $0.5 < z \le 1$ задает изоморфизм булевых формул и их решеточных продолжений.
7. Множество РПБФ является расширением множества булевых функций путем присоединения нестандартных конъюнкций вида $z_i \wedge \overline{z_i}$. Соответственно РПБФ разбивается на стандартную и нестандартную части. Нестандартная конъюнкция не превышает 0.5 на всем единичном кубе и равна 0 на всех вершинах.

Решеточное продолжение булевой функции позволяет различать, какой именно переменной или ее инверсией определяется значение функции. Назовем *конусом, ассоциированным с вершиной T* куба, множество точек куба, каждая координата которых отличается от соответствующей координаты вершины T меньше, чем на 0.5. Таким образом, соответствие конус-вершина является взаимно однозначным. *Удалением r* РПБФ (от значения 0.5) будем называть число аргументов, которые в метрике $\Delta_{ab} = |a - b|$ ближе к 0.5, чем значение функции.

Имеет место следующая лемма.

Лемма. РПБФ с конечным числом композиций представима в виде дизъюнктивной нормальной формы.

Следствие. Решеточно продолженная целевая функция имеет вид

$$H = C \vee g_1 z_1 \overline{z_1} \vee \ldots \vee g_N z_N \overline{z_N} \qquad (6)$$

где C — целевая конъюнкция (стандартная часть РПБФ), представленная выражением (3), g_i — решеточные продолжения булевых функций, при этом g_i не зависит от z_i.

Заметим, что поскольку каждое из преобразований шифрования может быть быстро вычислено на ЭВМ, описывающие это преобразование булевы формулы и их решеточные продолжения быстро вычислимы. Поэтому быстро вычисляется и решеточно продолженная целевая функция.

Если целевая функция содержит только стандартную конъюнкцию, то ключ может быть легко вычислен следующим алгоритмом. Придадим

переменным произвольные значения с попарно различными удалениями и вычислим функцию. Поскольку значение функции превышает 0.5 в единственном конусе, а вероятность угадать этот конус мала, будем считать, что значение функции меньше 0.5. Пусть значение функции равно некоторому разряду ключа (или его инверсии). Поскольку $H < 0.5$, то этот разряд был выбран неверно. Заменяем его на инверсный и повторно вычисляем функцию. Так повторяем до тех пор, пока значение целевой функции не превысит 0.5. Решением является вершина куба, ассоциированная с конусом, в котором $H > 0.5$. Алгоритм требует в среднем $N/2$ шагов.

3. Метод криптоанализа

Реально решеточно продолженная целевая функция содержит и нестандартные конъюнкции, в этом случае $C \le H$. Поэтому указанный выше метод нужно модифицировать с учетом последнего неравенства.

Пусть для набора значений переменных $\{Z_i\}$ ($Z_i \neq 0.5$ — значение переменной z_i) из конуса, ассоциированного с некоторой вершиной куба, удаление целевой РПБФ равно $r(H) = r(Z_k)$. При этом во всех конусах, кроме конуса, ассоциированного с решением, $H < 0.5$. Поскольку $C \le H$, то не все значения переменных с удалением не менее $r(Z_k)$ выбраны правильно. Чем больше удаление целевой функции, тем больше информации о ключе можно получить.

Целевая РПБФ может вычисляться как в разных конусах, так и в одном конусе с различным упорядочением переменных. После серии таких испытаний получается набор условий типа “хоть одна переменная из данного набора определена неправильно”. Если Z_k не является ближайшим к 0.5 числом из $\{Z_i\}$, то это позволяет отбраковать сразу значительную долю ключей, которые не могут удовлетворять целевой конъюнкции. Соответственно сужается область перебора.

Заметим, что если удастся доказать, что на данном наборе переменных $H = C$, то это позволяет раскрыть сразу все разряды ключа с удалением не меньшим, чем удаление целевой функции.

Криптоанализ может проводиться следующим алгоритмом.

Алгоритм.

1. По известным блокам шифрограммы и разрядам открытого текста составляется целевая функция.
2. Разрядам ключа придаются произвольные значения из случайно выбранного конуса с попарно различными удалениями. Вычисляется целевая функция. Если она определяется наименее удаленным разрядом ключа, то выбор конуса или выбор упорядочивания разрядов ключа неудачен. В противном случае вычисляется множество возможных ключей.

3. Предыдущий пункт повторяется до тех пор, пока не будет найден ключ, который вычисляется как пересечение множеств возможных ключей на каждой итерации п. 2.

Для повышения эффективности решеточного анализа можно использовать следующее:

1. Приравнивать не вычисленные значения разрядов шифрограммы и их истинные значения, а значения промежуточных текстов, которые получаются зашифрованием открытого текста на половине циклов и расшифрования шифрограммы на половине циклов на том же ключе (для этого нужно знать не отдельные разряды, а целые блоки открытых текстов).
2. Число отбракованных ключей для предложенного алгоритма равно $2^{r(H)}$. Поэтому желательно использовать не случайный выбор конуса и упорядочивания, а проводить процедуры, обеспечивающие максимум удаления.
3. Для снижения сложности метода целесообразно учитывать закон распределения вероятностей разностей удалений целевой функции и целевой конъюнкции.
4. На сложность метода сильно влияют нестандартные конъюнкции. С целью ослабления их действия булевы формулы, описывающие стандартную часть каждой операции шифрования, должны быть минимизированы.

Возможно, для каждого криптоалгоритма существует пара чисел, задающих оптимальные число перестановок значений разрядов ключа для каждого конуса и число конусов, которая позволяет минимизировать сложность раскрытия ключа. Поиск максимального удаления может выполняться указанным выше алгоритмом, если в п. 2 для каждого конуса выбирать оптимальное число перестановок и в качестве удаления целевой функции в данном конусе выбирать наибольшее.

Данный метод не работает, если в уравнении (6) почти все g_i больше 0.5. Заметим, что практически с ростом числа циклов шифрования удаление целевой функции стремится к 0, т. е. значения g_i превышают 0.5.

Метод должен быть эффективен для тех шифров, в которых не каждый разряд ключа сцеплен с каждым разрядом шифртекста, например, для шифров с псевдослучайным выбором слов ключа.

Литература

1. M. Matsui. Linear cryptanalysis method for DES cipher // *Advances in Cryptology* — EUROCRYPT ’93, LNCS, v. 765, 1994, pp. 386–397.
2. E. Biham, A. Shamir. Differential cryptanalysis of DES-like cryptosystems // *Advances in Cryptology* — CRYPTO ’90, LNCS, v. 537, 1991, pp. 2–21.