

Криптосистема с открытым ключом на основе задачи об F -выполнимости булевых формул

Е. А. Поцелуевская

В современном мире значительная часть информации обрабатывается в электронном виде. В связи с необходимостью обеспечить защиту такой информации при передаче по открытым каналам связи, широкое распространение получили криптографические системы с открытым ключом, основанные на различных NP-полных задачах. В настоящей работе рассматривается реализация асимметричной криптографической системы на основе NP-полной задачи об F -выполнимости булевых формул.

Ключевые слова: криптосистема, шифрование, выполнимость, булевы функции.

1. Основные понятия и утверждения

Задача об F -выполнимости булевых формул ставится следующим образом.

Пусть $\mathbf{F} = \{F_1, \dots, F_s\}$ — любое конечное множество формул (функциональных символов). Определим F -формулу как конъюнкцию $F_{i_1}(\cdot)F_{i_2}(\cdot) \dots F_{i_t}(\cdot)$ с переменными x_1, \dots, x_n , расставленными некоторым образом.

Проблема F -выполнимости — это проблема выполнимости F -формулы. В общем случае данная задача является NP-полной.

Пример 1. Пусть $F(x, y, z)$ — булева формула с таблицей истинности $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Тогда формула $F(x, y, z)F(x, y, u)F(u, u, y)$ выполнима и $(x, y, z, u) = (0, 1, 0, 0)$ — ее выполняющий набор.

Пусть дана система формул $F = \{F_1, \dots, F_s\}$. Для переменной x будем обозначать $x \in F_i$, если x входит в запись формулы F_i и является существенной переменной. Обозначим множества: $D(x) = \{F_i | x \in F_i\}$, для множества переменных M из $\{x_1, \dots, x_n\}$: $D(M) = \{F_i | \exists x \in M \quad x \in F_i\}$.

Определение. Будем говорить, что множество переменных M покрывает множество формул $F = \{F_1, \dots, F_s\}$, если $F \setminus D(M) = \emptyset$.

В случае, если F -формула задана в конъюнктивной нормальной форме (КНФ) и каждая из функций F_i зависит не более чем от трех переменных, для решения задачи F -выполнимости существует алгоритм, приведенный в статье [1]. Алгоритм основан на переборе минимального подмножества S переменных x_i , которые покрывают все дизъюнкции от трех переменных, входящие в КНФ, и решении для каждого фиксированного набора значений переменных из S полиномиальной подзадачи о 2-выполнимости.

Сложность данного алгоритма составляет $\left(1 + \sum_{i=1}^k 2^{|S_i|}\right) poly(|x|)$,

где $|x|$ — длина входа, множества S_i — это множества переменных, вычисляемые в ходе работы алгоритма, для которых выполнено:

$S = \bigsqcup_{i=1}^k S_i$. В случае если для всех $i = 1 \dots k$ $|S_i| \leq \log_2(poly(|x|))$

сложность алгоритма будет полиномиальной величиной.

Пусть теперь задано отображение $F : E_n \rightarrow E_n$:

$$F = \begin{cases} F_1(x_1, \dots, x_n); \\ F_2(x_1, \dots, x_n); \\ \dots \\ F_n(x_1, \dots, x_n). \end{cases}$$

Тогда справедлива следующая теорема.

Теорема 1 (критерий Хаффмана). Система функций (F_1, \dots, F_n) определяет подстановку тогда и только тогда, когда функции обладают следующим распределением весов:

$$\begin{cases} \|F_i\| = 2^{n-1}, \quad i \in \{1, \dots, n\}; \\ \|F_i F_j\| = 2^{n-2}, \quad i, j \in \{1, \dots, n\}; \quad j \neq i; \\ \dots \\ \|F_1 \dots F_n\| = 2^{n-n} = 1. \end{cases}$$

2. Формирование ключевой пары

Основным параметром криптографической системы с открытым ключом на основе задачи об F -выполнимости служит количество различных переменных, задействованных в F -формуле (обозначим его n).

Для формирования ключевой пары рассмотрим n булевых функций F_1, \dots, F_n , каждая из которых зависит не более чем от трех переменных, со следующими условиями (*):

- Для функций F_1, \dots, F_n выполнены условия Критерия Хаффмана:

$$\begin{cases} \|F_i\| = 2^{n-1}, i \in \{1, \dots, n\}; \\ \|F_i F_j\| = 2^{n-2}, i, j \in \{1, \dots, n\}; j \neq i; \\ \dots \\ \|F_1 \dots F_n\| = 1. \end{cases}$$

- Переменные x_1, \dots, x_n расставлены в формулах F_1, \dots, F_n таким образом, что минимальное количество переменных, покрывающих все формулы $m \leq \log_2(n)$.

Функции, удовлетворяющие данным условиям, существуют, как показано ниже.

Пример 2. Пусть количество переменных $n = 4$, тогда следующие функции удовлетворяют заданным условиям:

$$\begin{cases} F_1 = x_1 \oplus x_2 \oplus x_3; \\ F_2 = x_2 \oplus x_3 \oplus x_4; \\ F_3 = x_2 \oplus x_3; \\ F_4 = x_3 \oplus x_4 \oplus 1. \end{cases}$$

Минимальное множество переменных, покрывающих все формулы $\{x_3\}$, его мощность $m = 1 < \log_2(n)$.

Утверждение 1. Все функции F , зависящих не более чем от трех переменных, для которых выполнено условие $\|F\| = 2^{n-1}$, могут быть записаны полиномом Жегалкина степени не выше 2.

Доказательство. Так как функция F зависит от трех переменных, то в общем случае степень соответствующего полинома Жегалкина не может превышать 3. Докажем, что на самом деле степень не может быть равна 3.

Допустим, F может быть записана в виде полинома Жегалкина третьей степени, тогда F имеет вид:

$$F(x_1, x_2, x_3) = x_1x_2x_3 \oplus p_2(x_1, x_2, x_3),$$

где $p_2(x_1, x_2, x_3) = c_{1,2}x_1x_2 \oplus c_{1,3}x_1x_3 \oplus c_{2,3}x_2x_3 \oplus c_1x_1 \oplus c_2x_2 \oplus c_3x_3 \oplus c_0$ — полином Жегалкина степени не выше 2.

Тогда таблица истинности для $p_2(x_1, x_2, x_3)$ выглядит следующим образом:

(x_1, x_2, x_3)	$p_2(x_1, x_2, x_3)$
000	c_0
001	$c_3 \oplus c_0$
010	$c_2 \oplus c_0$
011	$c_{2,3} \oplus c_2 \oplus c_3 \oplus c_0$
100	$c_1 \oplus c_0$
101	$c_{1,3} \oplus c_1 \oplus c_3 \oplus c_0$
110	$c_{1,2} \oplus c_1 \oplus c_2 \oplus c_0$
111	$c_{1,2} \oplus c_{1,3} \oplus c_{2,3} \oplus c_1 \oplus c_2 \oplus c_3 \oplus c_0$

Тогда возможны следующие два варианта:

- 1) $p_2(1, 1, 1) = 0$. Тогда, так как $x_1x_2x_3$ принимает значение 1 только на наборе 111, а $\|F\| = 2^{n-1}$, для p_2 должно быть выполнено: $\|p_2\| = 2^{n-1} - 1$.
- 2) $p_2(1, 1, 1) = 1$, тогда для p_2 должно быть выполнено: $\|p_2\| = 2^{n-1} + 1$.

Таким образом, полином p_2 в любом случае должен принимать значение 1 на нечетном количестве наборов. Суммируя значения p_2 на всех наборах, получим:

$$c_0 \oplus (c_3 \oplus c_0) \oplus (c_2 \oplus c_0) \oplus (c_{2,3} \oplus c_2 \oplus c_3 \oplus c_0) \oplus (c_1 \oplus c_0) \oplus (c_{1,3} \oplus c_1 \oplus c_3 \oplus c_0) \oplus (c_{1,2} \oplus c_1 \oplus c_2 \oplus c_0) \oplus (c_{1,2} \oplus c_{1,3} \oplus c_{2,3} \oplus c_1 \oplus c_2 \oplus c_3 \oplus c_0) = 0.$$

Таким образом, число наборов, на которых p_2 принимает значение 1 всегда четно. Получили противоречие с исходным предположением, следовательно, степень F не может превышать 2. Утверждение доказано.

Каждую из выбранных функций F_1, \dots, F_n запишем в форме полинома Жегалкина. В соответствии с утверждением 1, данные полиномы имеют степень не выше 2.

Рассмотрим замену переменных $y = Ax + b$, где $y = (y_1, \dots, y_n)^T$, $x = (x_1, \dots, x_n)^T$ — столбцы переменных, A — невырожденная матрица размера $n \times n$, $a_{i,j} \in \{0, 1\}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, n\}$, $b = (b_1, \dots, b_n)^T$, $b_i \in \{0, 1\}$, $i \in \{1, \dots, n\}$. Расширенная матрица этой системы $P = (A|b)$ размера $n \times (n+1)$ будет служить закрытым ключом криптографической системы. Длина закрытого ключа равна $n^2 + n$ бит. Таким образом, злоумышленнику для подбора закрытого ключа потребуется $2^{n(n+1)}$ операций.

Заменив переменные по формуле $x = A^{-1}y + A^{-1}b$ и приведя подобные слагаемые, получим следующую систему формул:

$$\begin{cases} f_1(y_1, \dots, y_n); \\ f_2(y_1, \dots, y_n); \\ \dots \\ f_n(y_1, \dots, y_n). \end{cases}$$

Функции f_1, \dots, f_n представляют собой полиномы Жегалкина степени не выше 2 от переменных y_1, \dots, y_n . Далее рассмотрим следующее преобразование:

$$\begin{cases} g_1(Y_1, Y_2) = f_1(Y_1) \oplus f_2(Y_2); \\ g_2(Y_1, Y_2) = f_2(Y_1) \oplus f_3(Y_2); \\ \dots \\ g_n(Y_1, Y_2) = f_n(Y_1) \oplus f_1(Y_2). \end{cases}$$

При этом $\overline{Y_1} = (y_1^1, y_2^1, \dots, y_n^1)$, $\overline{Y_2} = (y_1^2, y_2^2, \dots, y_n^2)$ — два набора переменных. Для функций g_1, \dots, g_n осуществим линейное преобразование с помощью невырожденной матрицы A :

$$(h_1, h_2, \dots, h_n)^T = A(g_1, g_2, \dots, g_n)^T.$$

Функции h_1, \dots, h_n представляют собой полиномы Жегалкина степени не выше 2 от переменных $y_1^1, \dots, y_n^1, y_1^2, \dots, y_n^2$. Каждая из функций h_1, \dots, h_n однозначно задается строкой коэффициентов длины $2C_n^2 + 2C_n^1 + 1$, где C_n^k — число сочетаний из n по k . Стока коэффициентов всех функций h_1, \dots, h_n , упорядоченных по сначала по номеру функции, затем по степени, затем лексикографически, является открытым ключом системы. Длина открытого ключа составляет $n(2C_n^2 + 2C_n^1 + 1) = n(n^2 + n + 1)$.

3. Алгоритм шифрования

Вход алгоритма: M — открытый текст, K_B — открытый ключ получателя, n — количество переменных (параметр безопасности, фиксированный для данной криптосистемы).

Выход алгоритма: Зашифрованный текст C .

Шифрование осуществляется в следующем порядке

- 1) Исходное сообщение M разбивается на блоки длины n^2 , каждый из блоков разбивается на n подблоков длины n :

$$\underbrace{m_1, \dots, m_n}_{M_1} \underbrace{m_{n+1}, \dots, m_{2n}}_{M_2} \dots \underbrace{m_{n^2-n+1}, \dots, m_{n^2}}_{M_n}$$

- 2) Для каждого из блоков $M_1 \dots M_n$ блокам переменных Y_1, \dots, Y_n присваиваются соответствующие значения ($Y_1 = M_1, \dots, Y_n = M_n$) и подставляются в систему формул K_B следующим образом:

$$\left\{ \begin{array}{l} h_1(M_1, M_2) = c_1^1; \\ h_2(M_1, M_2) = c_2^1; \\ \dots \\ h_n(M_1, M_2) = c_n^1; \\ h_1(M_2, M_3) = c_1^2; \\ h_2(M_2, M_3) = c_2^2; \\ \dots \\ h_n(M_2, M_3) = c_n^2; \\ \dots \\ h_1(M_n, M_1) = c_1^n; \\ h_2(M_n, M_1) = c_2^n; \\ \dots \\ h_n(M_n, M_1) = c_n^n; \end{array} \right.$$

В результате каждый блок исходного текста длины n^2 преобразуется в блок шифротекста длины n^2 .

Операция шифрования однозначна: так как исходные функции удовлетворяют критерию Хаффмана, а линейные преобразования, использованные при формировании открытого ключа осуществляются с помощью невырожденной матрицы, то преобразование $M \rightarrow C$ является взаимно однозначным, и открытому тексту M при заданном ключе K_B соответствует ровно один шифротекст C .

4. Алгоритм расшифрования

Вход алгоритма: C — шифротекст, K_B — открытый ключ получателя, P_B — закрытый ключ получателя, n — количество переменных (параметр безопасности, фиксированный для данной криптосистемы).

Выход алгоритма: Исходный текст M .

Расшифрование осуществляется в следующем порядке

- 1) Зашифрованное сообщение C разбивается на блоки длины n^2 ;
- 2) Для каждого из блоков составляется система из n^2 уравнений:

$$\left\{ \begin{array}{l} h_1(Y_1, Y_2) = c_1^1; \\ h_2(Y_1, Y_2) = c_2^1; \\ \dots \\ h_n(Y_1, Y_2) = c_n^1; \\ h_1(Y_2, Y_3) = c_1^2; \\ h_2(Y_2, Y_3) = c_2^2; \\ \dots \\ h_n(Y_2, Y_3) = c_n^2; \\ \dots \\ h_1(Y_n, Y_1) = c_1^n; \\ h_2(Y_n, Y_1) = c_2^n; \\ \dots \\ h_n(Y_n, Y_1) = c_n^n. \end{array} \right.$$

- 3) Так как $P_B: y = Ax + b$, где $P_B = (A|b)$, A — невырожденная матрица размера $n \times n$, b — столбец высоты n , получатель может перейти к системе уравнений:

$$\left\{ \begin{array}{l} A(g_1(Y_1, Y_2), g_2(Y_1, Y_2), \dots, g_n(Y_1, Y_2))^T = C_1^T, \\ \qquad \qquad \qquad C_1 = (c_1^1, c_2^1, \dots, c_n^1); \\ A(g_1(Y_2, Y_3), g_2(Y_2, Y_3), \dots, g_n(Y_2, Y_3))^T = C_2^T, \\ \qquad \qquad \qquad C_2 = (c_1^2, c_2^2, \dots, c_n^2); \\ \dots \\ A(g_1(Y_n, Y_1), g_2(Y_n, Y_1), \dots, g_n(Y_n, Y_1))^T = C_n^T, \\ \qquad \qquad \qquad C_n = (c_1^n, c_2^n, \dots, c_n^n). \end{array} \right.$$

Решив каждую из n систем от n переменных g_1, \dots, g_n , получатель находит значения функций $f_i(Y_j)$ из следующей системы:

$$\left\{ \begin{array}{l} g_1(Y_1, Y_2) = f_1(Y_1) \oplus f_2(Y_2); \\ g_2(Y_1, Y_2) = f_2(Y_1) \oplus f_3(Y_2); \\ \dots \\ g_n(Y_1, Y_2) = f_n(Y_1) \oplus f_1(Y_2); \\ g_1(Y_2, Y_3) = f_1(Y_2) \oplus f_2(Y_3); \\ g_2(Y_2, Y_3) = f_2(Y_2) \oplus f_3(Y_3); \\ \dots \\ g_n(Y_2, Y_3) = f_n(Y_2) \oplus f_1(Y_3); \\ \dots \\ g_1(Y_n, Y_1) = f_1(Y_n) \oplus f_2(Y_1); \\ g_2(Y_n, Y_1) = f_2(Y_n) \oplus f_3(Y_1); \\ \dots \\ g_n(Y_n, Y_1) = f_n(Y_n) \oplus f_1(Y_1). \end{array} \right.$$

Таким образом, получатель приходит к следующей системе уравнений:

$$\left\{ \begin{array}{l} f_1(Y_1) = d_1^1; \\ f_2(Y_1) = d_2^1; \\ \dots \\ f_n(Y_1) = d_n^1; \\ \dots \\ f_1(Y_n) = d_1^n; \\ f_2(Y_n) = d_2^n; \\ \dots \\ f_n(Y_n) = d_n^n. \end{array} \right.$$

- 4) В формулах полученной системы f_1, \dots, f_n осуществляется замена переменных посредством матрицы P_B : $y = Ax + b$, где $P_B = (A|b)$, A — матрица размера $n \times n$, b — столбец высоты n . В результате данного шага для каждого из подблоков длины n должна быть получена исходная система формул, зависящих от переменных x_1, \dots, x_n :

$$\begin{cases} F_1(x_{1,1}, x_{1,2}, x_{1,3}), & x_{1,i} \in \{x_1, \dots, x_n\}, i \in \{1, 2, 3\}; \\ F_2(x_{2,1}, x_{2,2}, x_{2,3}), & x_{2,i} \in \{x_1, \dots, x_n\}, i \in \{1, 2, 3\}; \\ \dots \\ F_n(x_{n,1}, x_{n,2}, x_{n,3}), & x_{n,i} \in \{x_1, \dots, x_n\}, i \in \{1, 2, 3\}. \end{cases}$$

Таким образом, получим n систем уравнений вида:

$$\begin{cases} F_1 = d_1; \\ \dots \\ F_n = d_n. \end{cases}$$

- 5) Составляется F -формула следующим образом: в случае, если известно, что функция F_i принимает значение 1, то она входит в формулу без отрицания; в противном случае — с отрицанием. То есть F -формула имеет вид: $F_1^{d_1} F_2^{d_2} \dots F_n^{d_n}$.
- 6) Отрицания вносятся в выражения полинома Жегалкина для каждой из функций:

$$\underbrace{(F_1 \oplus d_1 \oplus 1)}_{F'_1} \underbrace{(F_2 \oplus d_2 \oplus 1)}_{F'_2} \dots \underbrace{(F_n \oplus d_n \oplus 1)}_{F'_n}.$$

- 7) Полученные функции F'_1, \dots, F'_n представляются в виде КНФ. Так как все функции F'_1, \dots, F'_n зависят не более, чем от трех переменных и для них выполнены условия (*), то общее количество таких функций $C_8^4 = 70$. Поэтому для преобразования формул к КНФ достаточно постоянно хранить в памяти матрицу соответствия полиномов Жегалкина и КНФ для всех 70 функций.
- 8) Для формулы $F'_1 \& \dots \& F'_n$, заданной в КНФ, решается задача об F -выполнимости в соответствии с алгоритмом [1], на выходе для каждого из n подблоков имеем выполняющий набор $(x_1^i, \dots, x_n^i) = (\sigma_1, \dots, \sigma_n)$, где $i \in \{1, \dots, n\}$ — номер подблока.

- 9) Набор (x_1^i, \dots, x_n^i) преобразуется по формуле $X^i = A^{-1}Y^i + A^{-1}b$, в итоге получаем $Y^i = (y_1^i, \dots, y_n^i) = (m_1^i, \dots, m_n^i)$ — блок исходного текста M длины n .

Операция расшифрования однозначна: так как исходные функции удовлетворяют критерию Хаффмана, а линейные преобразования, использованные при формировании открытого ключа осуществляются с помощью невырожденной матрицы, то преобразование $C \rightarrow M$ является взаимно однозначным, и зашифрованному тексту C при заданной ключевой паре (P_B, K_B) соответствует ровно один исходный текст M .

5. Оценка сложности работы криптосистемы

Очевидно, что на этапе шифрования сложность работы алгоритма линейна относительно длины входной информации n^2 (для каждого из блоков).

Оценка сложности работы алгоритма расшифрования приведена ниже.

Утверждение 2. Пусть ключевая пара (P_B, K_B) выбрана в соответствии с условиями (*). Тогда для каждого из блоков шифротекста сложность работы алгоритма расшифрования составляет $poly(n)$.

Доказательство. Шаги 1–3 выполняются за полиномиальное число шагов, так как решение n систем линейных уравнений методом Гаусса требует $O(n^4)$ операций. Шаг 4 алгоритма требует полиномиального числа операций, так как формулы замены переменных линейны. Шаги 5–7 алгоритма требуют линейного числа операций (очевидно). На шаге 8 сложность алгоритма, учитывая выполнение второго условия (*) оценивается как $\left(1 + \sum_{i=1}^k 2^{|S_i|}\right) poly(n) \leq (1 + 2^{|S^1|}) poly(n) \leq (1 + 2^m) poly(n) \leq (1 + poly(n)) poly(n)$. Обратное преобразование $Y^i \rightarrow X^i$ требует полиномиального числа операций. Таким образом, сложность расшифрования полиномиальна. Утверждение доказано.

6. Оценка надежности криптосистемы

Для нахождения открытого текста M при заданном шифротексте C и открытом ключе K_B , злоумышленнику требуется решить задачу F -выполнимости для формулы $h_1(Y_1, Y_2)^{c_1} \& \dots \& h_n(Y_n, Y_1)^{c_n}$, которая NP-полна и сводится к перебору n^2 значений, либо определить исходные функции F_1, \dots, F_n . Однако без знания закрытого ключа P_B вычисление исходной системы функций также сводится к перебору. Таким образом, взлом криптографической системы потребует 2^{n^2} операций.

На текущий момент в используемых на практике криптосистемах с открытым ключом используется длина закрытого ключа в 256 бит. Для обеспечения аналогичного уровня криптостойкости в криптосистеме на основе задачи F -выполнимости достаточно задать параметр $n = 16$. В этом случае длины ключей будут следующими: $|P| = 272$, $|K| = 4368$.

Для сокращения длины открытого ключа может быть использована следующая модификация исходной криптосистемы. Пусть для преобразования переменных используется невырожденная матрица A_1 ($y = A_1x + b$), такая что каждая из переменных x_i ($i \in \{1, \dots, n\}$) зависит не более чем от двух переменных y_k ($k \in \{1, \dots, n\}$). Для преобразования функций (f_1, \dots, f_n) к функциям g_1, \dots, g_n используется невырожденная матрица A_2 , такая что каждая функция $g_i(Y_1, Y_2)$ зависит ровно от двух функций $f_m(Y_1), f_p(Y_2)$. Тогда закрытый ключ — это запись вида $(A_1|b|A_2)$, и его длина составляет $n(4 \log_2 n + 1)$.

Каждая функций из записи открытого ключа при заданных параметрах зависит от ровно двух наборов Y_k, Y_l , при этом в каждом из наборов задействовано не более 6 переменных. Тогда открытый ключ может быть записан $n(12 \log_2 n + 37)$ битами. При параметре $n = 16$ длина закрытого ключа составит 272 бита, длина открытого ключа — 1360 бит.

Автор работы выражает признательность В. А. Носову за научное руководство.

Список литературы

- [1] Поцелуевская Е. А. Полиномиальные случаи решения задачи об F -выполнимости булевых формул // Интеллектуальные системы. Т. 12, вып. 1–4. 2008. С. 351–362.
- [2] Алексеев В. Б., Носов В. А. NP-полные задачи и их полиномиальные варианты. Обзор // Обозрение прикладной и промышленной математики. Т. 4, вып. 2. 1997. С. 165–193.
- [3] Schaefer T. J. The complexity of satisfiability problems // Proceedings of the 10th ACM Symposium on Theory of Computing. 1978. P. 216–226.