Правительство Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"
Московский институт электроники и математики им. А.Н. Тихонова
Департамент прикладной математики

УДК ______
№ госрегистрации ______
Инв. №______

УТВЕРЖДАЮ

____
головной исполнитель

__________________________ ____________

«______» ____________________ 2022 г.

МЕЖДИСЦИПЛИНАРНАЯ КУРСОВАЯ РАБОТА

по теме:
Исследование вопросов оптимизации методов анализа некоторых схем шифрования сохраняющих формат
(промежуточный)

Руководитель курсовой работы __________________ Д.Б. Фомин

Академический руководитель
образовательной программы ____________________ А.Б. Лось

Москва 2022

СПИСОК ИСПОЛНИТЕЛЕЙ

Выполнил студент ________________ Щеглова П.Н.

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей работе применяются следующие термины с соответствующими определениями и сокращениями:

Обозначения и сокращения

с.в. — случайная величина ;

Определения

Функции

$$a \oplus b = \begin{cases} 0, (a,b) \in \{(0,0),(1,1)\} \\ 1, (a,b) \in \{(0,1),(1,0)\} \end{cases}$$

$$Ind(Expr) = \begin{cases} 1, Expr = True \\ 0, Expr = False \end{cases}$$

$$(\alpha, b) = (\alpha_1 \cdot b_1) \oplus ... \oplus (\alpha_N \cdot b_N); \alpha = [\alpha_1, ..., \alpha_N], b = [b_1, ..., b_N];$$

1 Линейный метод

1.1 Схема и обозначения

Сначала опишем общую схему алгоритма и обозначения для применения линейного метода криптоанализа.

Известно T пар открытых текстов и соответствующих шифртекстов $(a^{(i)}, c^{(i)}), i \in \overline{1,T}$, каждый из которых состоит из N бит: $a_1^{(i)}, ..., a_N^{(i)}$ и $c_1^{(i)}, ..., c_N^{(i)}$. Пусть схема шифропреобразования с ключом K разбита на две последовательные части F_{K_1} и F_{K_2}, как показано на рисунке 1.

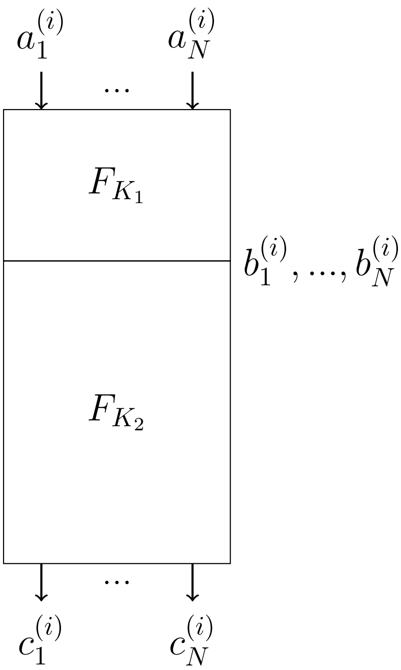

Рисунок 1 — Схема разбиения алгоритма на два блока для проведения линейного криптографического анализа

В первой из них используется часть исходного ключа K_1, во второй, соответсвенно, K_2 (при этом K_1 может частично совпадать с K_2). $F_{K_1'}(a^{(i)}) = b^{(i)} = b_1^{(i)}, ..., b_N^{(i)}$ – промежуточный шифртекст, зашифрованный на некотором ключе $K_1^{'}$. $\alpha = \alpha_1, ..., \alpha_N; \beta = \beta_1, ..., \beta_N$ – битовые маски, которые мы будем накладывать на промежуточный и итоговый шифртексты, соответственно. Наложение маски подразумевает скалярное произведение двух векторов: $(\alpha, b^{(i)})$.

Для отбраковывания ложных ключей линейный метод предполагает проверку выполнения некоторого соотношения с нужной вероятностью. Для двух масок $\alpha \in \mathbb{F}_2^n$ и $\beta \in \mathbb{F}_2^m$ и функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ определим следующую величину:

$$C_{\alpha,\beta}^F = 2 \cdot P\Big(\big(\alpha, x\big) = \big(\beta, F(x)\big), x \in \mathbb{F}_2^n\Big) - 1 =$$

$$2 \cdot \left(\frac{\sum_{x \in \mathbb{F}_2^n} \big(-1\big)^{\big(\alpha,x\big) \oplus \big(\beta,F(x)\big)}}{2 \cdot 2^n} + \frac{1}{2} \right) - 1 = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} \big(-1\big)^{\big(\alpha,x\big) \oplus \big(\beta,F(x)\big)}$$

и назовем ее преобладанием.

Для равномерно распределенной функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ справедлива следующая теорема:

1.2 Теорема ([1])

Пусть определено преобладание $C_{\alpha,\beta}^F$ для равномерно распределенной функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Тогда случайная величина $\xi = 2^{n-1}(C_{\alpha,\beta}^F + 1)$ *(у [1] речь идет о $Imb(\alpha, \beta) = 2^{n-1} \cdot C_{\alpha,\beta}^F$)* имеет биномиальное распределение $Bi\big(2^n, \dfrac{1}{2}\big)$ с математическим ожиданием $M\xi = 2^{n-1}$ и дисперсией $D\xi = 2^{n-2}$. В частности, при $n \to \infty$ распределение $2^{n/2}C_{\alpha,\beta}^F$ сходится к стандартному нормальному распределению $\mathcal{N}(0,1)$ *(об этом в [1] ничего нет).*

Осталось вывести переход к $C_{\alpha,\beta}^{F_1,\dots,F_r}$.

1.3 Алгоритм метода

Перейдем к описанию алгоритма. α и β заданы, вычислено теоритическое значение $C_{\alpha,\beta}^F$, вычислен доверительный интервал. Для каждого K_1':

а) Полагаем $\overline{P} = 0$;

б) Для каждого $a^{(i)}, i \in \overline{1,T}$, вычисляем $b^{(i)} = F_{K_1'}(a^{(i)})$;

в) Проверяем выполнено ли равентсво $(\alpha, b^{(i)}) = (\beta, c^{(i)})$.

г) Если равенство выполнилось, полагаем $\overline{P} = \overline{P} + 1$

д) После перебора материала полагаем $\overline{P} = \dfrac{\overline{P}}{T}$;

е) Если $\overline{P} \backsimeq P$, считаем, что часть ключа $K_1 = K_1'$, при необходимости продолжаем работу с F_{K_2} по той же схеме.

ж) Иначе, отбрасываем ключ K_1' как ложный, выбираем новый и повторяем все итерации.

Чем больше при этом T и $|C^F_{\alpha,\beta}|$, тем большая доля значений $K_1^{'}$ будет отбракована на каждой итерации, вплоть до однозначного определения $K_1^{'}$.

Для того, чтобы применить вычисляемую оценку для отбраковывания ложных ключей, необходим различитель, который на основе теоритической $C^F_{\alpha,\beta}$ определяет, выполнилось ли соотношение с нужной вероятностью. Чтобы построить различитель, воспользуемся результатами, полученными в [2].

2 Эксперименты

Оценивание значения преобладания $C^F_{\alpha,\beta}$ позволяет оценить и эффективность линейного метода. Обычно вместо непосредственно преобладания оценивают величину $(C^F_{\alpha,\beta})^2$. В качестве оценки указанной случайной величины используем статистику:

$$S = \left(\frac{2}{T} \cdot \sum_{i=1}^{T} v_i - 1\right)^2$$

где T - количество материала, $v_i = Ind\Big(\big(\alpha, x_i\big) = \big(\beta, F(x_i)\big)\Big)$ - реализация независимых случайных величин, распределенных по биномиальному закону с вероятностью 0 равной $\dfrac{C^F_{\alpha,\beta}+1}{2}$, а $x_i, F(x_i)$ – i-ые текст и соответствующий шифртекст. Характеристики статистики S зависят от функции F, поэтому рассмотрим различные случаи и проведем для них эксперименты, чтобы подтвердить корректность теоритических вычислений.

2.1 Эксперимент № 1

Целью первого типа экспериментов является рассмотрение статистики S в случае, когда функция F - биективное отображение (перестановка на множестве текстов) $\mathbb{F}_2^n \to \mathbb{F}_2^n$. $v_i = Ind\Big(\big(\alpha, x_i\big) = \big(\beta, F(x_i)\big)\Big)$, $\overline{v} = (v_1, v_2, ..., v_{2^n})$ – вектор из $\mathbb{F}_2^{2^n}$, для дальнейшего использования полученных значений при отсеивании ложных ключей.

Возьмем случайную величину $\xi = \sum_{i=1}^{2^n} v_i, \xi \in \{0, 1, ..., 2^n\}$, соответствующую количеству единиц в векторе $\overline{v}$ в зависимости от истинной подстановки. Тогда всего существует $\dbinom{2^n}{\xi}$ возможных векторов, для которых количество единиц совпадает с истинным.

При применении линейного метода криптоанализа проверяются лишь первые T координат вектора $\overline{v}$, число T соответствует количеству материала, т.е. количеству известных пар открытого и шифрованного текстов. В таком случае, при наблюдениях происходит переход к случайной величине $\xi_T = \sum_{i=1}^{T} v_i, \xi_T \in \{0, 1, ..., T\}$. Найдем математическое ожидание для с.в. $\eta = \varphi(\xi_T)$, где φ - произвольная функция определенная на множестве целых чисел $\mathbb{Z}$:

$$E\eta = E\varphi(\xi_T) = \sum_{j=0}^{T} \varphi(j) \cdot P(\xi_T = j)$$

Вероятность события $\xi_T = j$ можно представить в виде суммы вероятностей с помощью формулы Байеса, где гипотезы $\{\xi = k\}_{k=\overline{0,2^n}}$ образуют полную группу событий:

$$E\eta = \sum_{j=0}^{T} \varphi(j) \cdot P(\xi_T = j) = \sum_{j=0}^{T} \varphi(j) \cdot \sum_{k=0}^{2^n} P(\xi_T = j | \xi = k) P(\xi = k)$$

Условную вероятность $P(\xi_T = j | \xi = k)$ можно подсчитать по классической вероятностной схеме, так как вероятности значений координат в векторах $\overline{v}$ идентичны для случайных величин ξ и ξ_T: всего вариантов выбрать координаты вектора, значение которых 1, равно $\begin{pmatrix} 2^n \\ k \end{pmatrix}$, число вариантов выбрать при этом ровно j первых T координат: $\begin{pmatrix} T \\ j \end{pmatrix} \begin{pmatrix} 2^n - T \\ k - j \end{pmatrix}$. Соответственно:

$$E\eta = \sum_{j=0}^{T} \varphi(j) \cdot \sum_{k=0}^{2^n} P(\xi = k) \frac{\begin{pmatrix} T \\ j \end{pmatrix} \begin{pmatrix} 2^n - T \\ k - j \end{pmatrix}}{\begin{pmatrix} 2^n \\ k \end{pmatrix}}$$

В случае когда $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ - случайная и равновероятная (т.е. выбор образа для заданного элемента множества открытых текстов считаем сделанным по равновероятной схеме), $\big(\alpha, x_i\big)$ и $\big(\beta, F(x_i)\big)$ также распредены по равновероятной схеме (наложение маски не влияет на общее распределение, так как ...did not get), а значит и $Ind\Big(\big(\alpha, x_i\big) = \big(\beta, F(x_i)\big)\Big)$.

Таком образом, получаем, что с.в. $\xi = \sum_{i=1}^{2^n} v_i$ распределена по биномиальному закону – $Bi\left(2^n, \dfrac{1}{2}\right)$ и

$$P(\xi = k) = \binom{2^n}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{2^n-k} = \frac{1}{2^{2^n}}\binom{2^n}{k}.$$

В таком случае, математическое ожидание для с.в. $\eta = \varphi(\xi_T)$ принимает следующий вид:

$$E\eta = \sum_{j=0}^{T} \varphi(j) \cdot \sum_{k=0}^{2^n} \frac{1}{2^{2^n}} \binom{2^n}{k} \frac{\binom{T}{j}\binom{2^n-T}{k-j}}{\binom{2^n}{k}} =$$

$$= \frac{1}{2^{2^n}} \sum_{j=0}^{T} \varphi(j) \cdot \binom{T}{j} \sum_{k=0}^{2^n} \binom{2^n-T}{k-j}$$

С помощью выведенной формулы получаем матожидание статистики $\big(\varphi(\xi_T) = (2\xi_T/T - 1)^2\big)$:

$$E\left(\frac{2}{T} \cdot \sum_{i=1}^{T} v_i - 1\right)^2 = \frac{1}{2^{2^n}} \sum_{j=0}^{T} \left(\frac{2j}{T} - 1\right)^2 \cdot \binom{T}{j} \sum_{k=0}^{2^n} \binom{2^n-T}{k-j}$$

Теперь посчитаем его для различных небольших n и T:

n	T	S
8	2^8	$0.00390625 = 2^{-8}$
8	2^7	$0.0078125 = 2^{-7}$
8	2^6	$0.015625 = 2^{-6}$
12	2^{12}	$0.000244141 \approx 2^{-12}$
12	2^{11}	$0.000488281 \approx 2^{-11}$
12	2^{10}	$0.0009765625 \approx 2^{-10}$
24	2^{24}	$5.96046e - 08 \approx 2^{-24}$
24	2^{22}	$2.38419e - 07 \approx 2^{-22}$
24	2^{16}	$1.52588e - 05 \approx 2^{-16}$

Отметим, что статистика для таких функций зависит именно от объема материала T, в то время как для самой случайной величины значение преобладания зависим только от n.

2.1.1 Реализация эксперимента

Теперь проведем эксперимент: возьмем все возможные тексты размера $n = 8$, 12 и 24 бит, сгенерируем несколько случайных подстановок F (100-1000, для получения усредненных значений) и будем проводить шифрование, применяя получившуюся функцию на выбранном объеме материала. Фиксируем произвольную маску и для каждой F вычисляем статистику S (и на входе, и на выходе маска берется одна и та же) на заданном количестве материала.

Реализацию эксперимента можно найти по ссылке. Полученные результаты совпадают с вычисленными теоритически.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Daemen Joan, Rijmen Vincent. Probability distributions of correlation and differentials in block ciphers // Journal of Mathematical Cryptology. — 2007. — Vol. 1, no. 3. — P. 221–242. — Access mode: https://doi.org/10.1515/JMC.2007.011.

2. Beyne Tim. Linear Cryptanalysis of FF3-1 and FEA. — 2021. — Access mode: https://www.esat.kuleuven.be/cosic/publications/article-3384.pdf (online; accessed: 25.05.2022).