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

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

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

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

ВВЕДЕНИЕ

С ускорением глобальной информатизации все острее встает вопрос о защите информации, в частности персональных данных. Несмотря на то, что существуют законы, регламентирующие порядок хранения и обработки персональных данных, возлагающие ответственность за их сохранность на операторов персональных данных, в большинстве случаев эта информация хранятся в базах в открытом виде, и несанкционированный доступ к ней не требует больших усилий от злоумышленника. В связи с тем, что последствия реализации данного типа угроз могут быть достаточно серьезными, остро встает задача безопасного хранения подобных данных. Для персональной информации наиболее подходящим способ защиты является шифрование с сохранение формата (format-preserving encryption, FPE), так как в отличие от традиционных механизмов шифрования, оно, во-первых, позволяет программам, обрабатывающим данные как переменные заранее заданного типа, так же успешно обрабатывать и зашифрованные данные, и, во-вторых, позволяет скрыть сам факт шифрования. В 2021 году Тим Бейн, аспирант Лёвенсокго католического университета в Бельгии, представил работу [**?**], в которой продемонстрировал, как можно уменьшить сложность атак на FPE-алгоритмы с настройками с помощью линейного криптографического анализа. В данной курсовой работе демонстрируются: описание линейного метода анализа схем FPE с настройками на основе сети Фейстеля, а именно стандарта FEA-1; применение линейного метода с акцентом на использование статистических критериев с использованием теоретических обоснований, представленных в анализируемой статье; а также результаты эксперимента по нахождению линейного статистического аналога для входных и выходных последовательностей шифропреобразования.

1 Шифрование с сохранением формата

1.1 Описание концепции

Format-preserving encryption (FPE) — это семейство перестановок на произвольном множестве $\mathcal{S}$, индексируемое ключом K [**?**]

$$FPE_K : \mathcal{S} \to \mathcal{S}.$$

Примеры отображений: шифрование 16–значного номера банковской карты 16–значным числом; шифрование одного английского слова другим английским словом. Блочный шифр — частный случай FPE–схемы, для которой $\mathcal{S} = \{0,1\}^n$, где n — длина блока.

Истинно случайная перестановка является идеальным шифром FPE, однако для больших множеств невозможно предварительно сгенерировать и запомнить такую перестановку. Таким образом, проблема FPE состоит в том, чтобы сгенерировать псевдослучайную перестановку из секретного ключа так, чтобы время вычисления для одного значения было небольшим (в идеале постоянным, но, что наиболее важно, меньшим, чем $O(n)$, где n - размер входных данных).

Алгоритм FPE можно реализовать с использованием сети Фейстеля. Например, стандарты FF1 и FF3-1 [**?**] берут за основу алгоритма сеть Фейстеля, а в качестве раундовой функции шифрования части входных данных F стандартизированный блочный шифр с блоками длины 128 бит (AES).

1.2 Действующие стандарты

Существует множество реализованных алгоритмов типа FPE, к действующим можно отнести разработанные в США FF1 и FF3-1 [**?**], а также южно-корейские FEA-1 и FEA-2 [**?**]. Алгоритм FEA, представленный институтом исследований национальной безопасности (NSR), также основан на сети Фейстеля, аналогично стандартам NIST, FF1 и FF3-1.

Разница между FEA-1 и FEA-2 состоит в том, что FEA-1 имеет размер настройки (параметра, подающегося на вход раундовой функции F) $128 - n$ бит (где n - размер входной последовательности), каждый с 12, 14 и 16 раундами при длине двоичного ключа 128, 192 и 256, соответственно. FEA-2 имеет фиксированный размер настройки в 128 бит с 18, 21 и 24 раундами при длинах ключей 128, 192 и 256, соответственно.

1.2.1 FEA-1

Опишем подробнее стандарт, который анализируется в данной работе, а именно FEA-1:

На вход алгоритму подаются последовательности чисел из конечного множества, мощностью от 2^8 до 2^{128}, размер двоичного ключа K может составлять 128, 192 или 256 бит. Алгоритм представляет собой последовательное применение итераций сети Фейстеля, ее общая схема представлена в левой части рисунка 1. Входная последовательность X на каждом раунде делится на две равные части X_a и X_b, X_b передается на вход F-функции, общая схема которой обозначена в правой части рисунка 1: T_a и T_b - левая и правая половины настройки, принцип формирования которой будет описан далее, RK_a и RK_b - левая и правая половины раундового ключа, S - блок подстановки (в данной схеме применяются идентичные S-блоки), $\mathcal{M}$ - блок умножения на заданную матрицу.

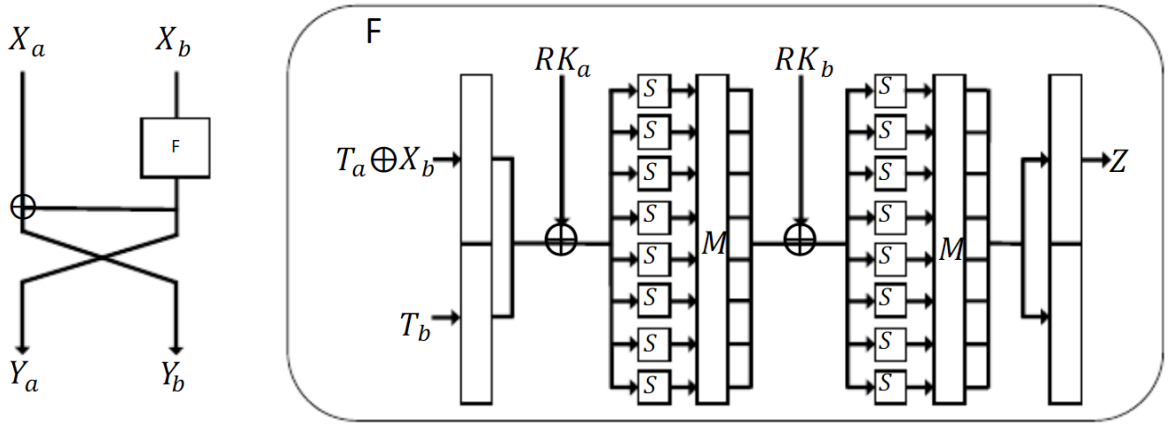

Рисунок 1 — Структура итерации FEA, на основе сети Фейстеля

Выбор настройки для каждого раунда происходит по следующему алгоритму: настройка T (битовый вектор длины $128 - n$) делится на две под-настройки $T_L = T_{[0:64-n_2-1]}$ и $T_R = T_{[64-n_2:128-n-1]}$ длины $64 - n_2$ и $64 - n_1$, соответственно. Полагаем $T_a^i = 0$ для каждой итерации и T_b^i для i-ой итерации, как:

$$T_b^i = \begin{cases} T_L & \frac{i}{2} \in N \\ T_R & \frac{i+1}{2} \in N \end{cases}$$

2 Линейный метод

2.1 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 2. *Нарисовать свой рисунок, заменить обозначения шифртекста и проме-*

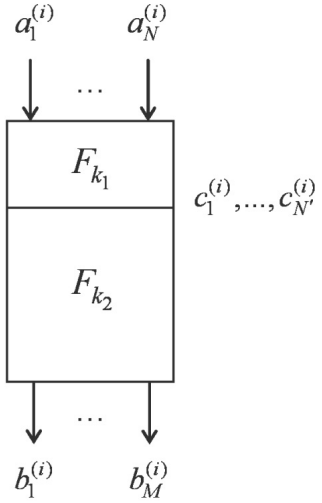

Рисунок 2 — Схема разбиения алгоритма на два блока для проведения линейного криптографического анализа

жуточного шифртекста

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:

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

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

и назовем ее преобладанием.

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

2.2 Теорема ([?])

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

Осталось вывести переход к $C^{F_1,\ldots,F_r}_{\alpha,\beta}$.

2.3 Алгоритм метода

Перейдем к описанию алгоритма. α и β заданы, вычислено теоритическое значение $C^F_{\alpha,\beta}$, 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 [**?**].

3 Эксперименты

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

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

где $v_i = Ind\Big(\big(\alpha, x_i\big) = \big(\beta, F(x_i)\big)\Big)$ - реализация независимых случайных величин, распределенных по биномиальному закону по теореме , $x_i, F(x_i)$ – i-ые открытый и шифрованный тексты, соответственно, а N - количество материала. При этом по построению предполагается, что вероятность

$$P(v_i = 0) = \frac{C^F_{\alpha,\beta} + 1}{2}.$$

3.1 Эксперимент № 1

Целью первого типа экспериментов является оценка $C^F_{\alpha,\beta}$ в случае, когда функция F - биективное отображение $\mathbb{F}_2^n \to \mathbb{F}_2^n$, причем выбор образа для заданного элемента множества открытых текстов считаем сделанным по равновероятной схеме. $v_i = Ind\Big(\big(\alpha, x_i\big) = \big(\beta, F(x_i)\big)\Big)$, $\overline{v} = (v_1, v_2, ..., v_{2^n})$ – вектор из $\mathbb{F}_2^{2^n}$.

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

При применении линейного метода криптоанализа проверяются лишь первые (для определенности) N координат вектора $\overline{v}$, число N соответствует количеству материала , т.е. количеству известных пар

открытого и шифрованного текстов. В таком случае возникает случайная величина $\xi_N = \sum_{i=1}^{N} v_i, \xi_N \in \overline{0, N}$. Найдем математическое ожидание для с.в. $\eta = \varphi(\xi_N)$, где φ - произвольная функция определенная на множестве целых чисел $\mathbb{Z}$:

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

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

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