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

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 [1], 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 [2]

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

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 [3] 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 [3], а также южно-корейские FEA-1 и FEA-2 [4]. 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.

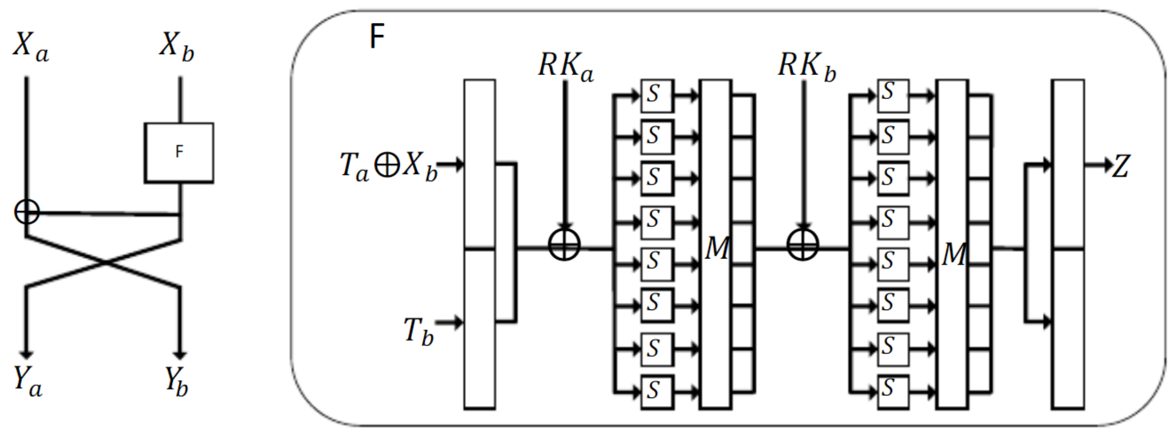

Рисунок 1 — Структура итерации FEA, на основе сети Фейстеля

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:

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

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. В первой из

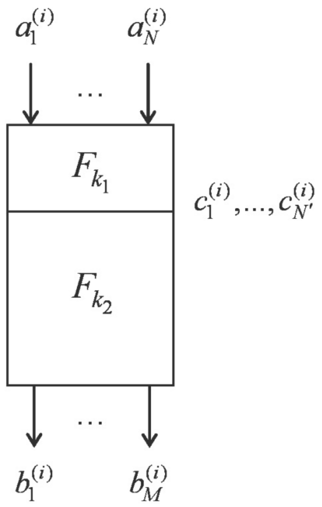

Рисунок 2 — Схема разбиения алгоритма на два блока для проведения линейного криптографического анализа

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: $(\alpha, b^{(i)}) = (\alpha_1 \cdot b_1^{(i)}) \oplus ... \oplus (\alpha_N \cdot b_N^{(i)})$.

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:

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

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

2.2 Теорема ([5])

Пусть определена $\delta^F_{\alpha,\beta}$ для равномерно распределенной функции $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Тогда случайная величина $\xi = 2^{n-1}(\delta^F_{\alpha,\beta}+1)$ имеет биномиальное распределение с математическим ожиданием $M\xi = 2^{n-1}$ и дисперсией $D\xi = 2^{n-2}$. В частности, при $n \to \infty$ распределение $2^{n/2}\delta^F_{\alpha,\beta}$ сходится к стандартному нормальному распределению $\mathcal{N}(0,1)$.

2.3 Алгоритм метода

Перейдем к описанию алгоритма. α и β заданы, вычислено теоритическое значение $\delta^F_{\alpha,\beta}$, 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 $\overline{P} \simeq P$, 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 $|\delta^F_{\alpha,\beta}|$, 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 δ определяет, выполнилось ли соотношение с нужной вероятностью. Чтобы построить различитель, воспользуемся результатами, полученными в [1].

2.4 Теоритические выкладки из [1]

Пусть $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ - наша раундовая функция, n - размер блока. Различитель строится на основе линейных соотношений с большой δ. Линейное соотношение для F определяется двумя масками $(L^{'}, L^{''}) \in \mathbb{F}_2^n \oplus \mathbb{F}_2^n$, и их δ равна:

$$\delta^F_{L^{'},L^{''}} = 2 \cdot P\left(L^{'}\&F(x) = L^{''}\&x\right) - 1,$$

где вероятность считается от равномерно распределенного открытого текста $x \in \mathbb{F}_2^n$, а & означает побитовое наложение маски на текст. Если $L^{'} \neq 0$, то математическое ожидание δ равновероятно распределенной функции равно нулю, а стандартное отклонение $\sigma = 2^{-n/2}$. Следовательно, если $\hat{\delta}$ значительно превосходит $2^{-n/2}$, то различителем можно считать вычисление $\hat{\delta}$ для возможных пар масок и сравнение получившего значения с некоторым заданным пороговым значением.

Основное наблюдение [1], которое можно эксплуатировать в атаках на FPE шифры, состоит в том, что шифр оказывается нестойким, если настройка (её часть) считается частью входных данных.

Рассмотрим два раунда FEA-1 (рисунок 3), настройка T_L - произвольная постоянная, а T_R считается переменной входа. Если это не так, то для проведения атаки T_R должна быть известной. Идея атаки состоит в том, что δ линейного соотношения раундовой функции F_r превышает $2^{-n/2}$ с достаточно большой вероятностью, что важно, когда настройка включается во входные данные, потому что область определения функции, которая отображает настройку и открытый текст в шифртекст, велика. Математическое ожидание δ линейных соотношений над случайной функцией с тем же размером входа (включая T_R размера $64 - n_1$), что и в FEA-1, равно нулю, а стандартное отклонение $2^{-32-n_1/2}$.

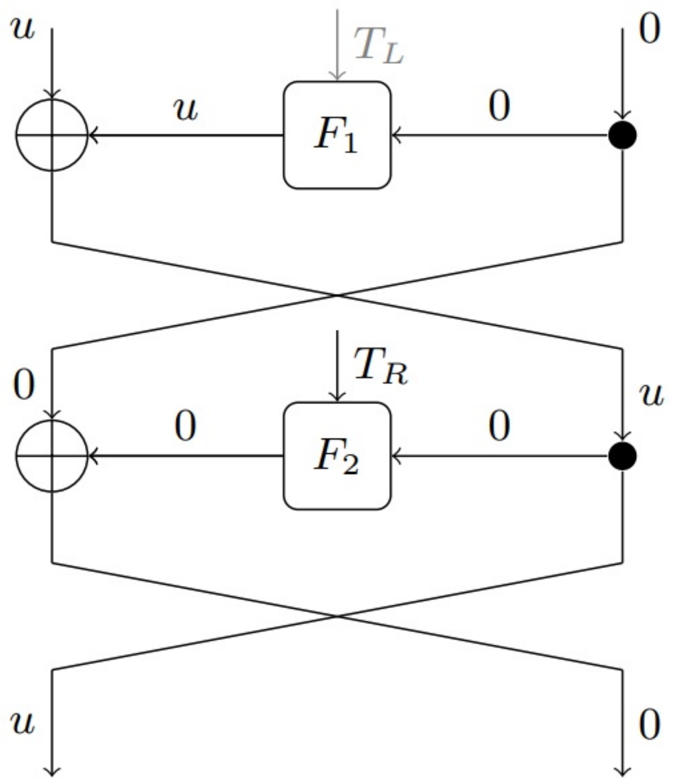

Рисунок 3 — Две итерации FEA-1

2.4.1 Применение

Пусть $r \geqslant 2$ - четное целое число. δ для r раундов равна $\delta = \prod_{i=1}^{r/2} \delta_i$, где $\delta_i \sim \mathcal{N}(0{,}1/\sqrt{N})$ по теореме 2.2, N — мощность множества текстов. Переменнные δ_i будем считать независимыми, что следует из предположения о независимости раундовых функций $F_1, F_3, ..., F_{r-1}$.

Эвристически вычислимо, что для FEA-1

$$1/\mathbb{E}(\delta^2) = \sqrt{N}^{r/2}, \tag{1}$$

где $\mathbb{E}(\delta^2)$ - средне-квадратичная δ для равномерно распределенного случайного ключа.

3 Эксперимент

Теперь проведем эксперимент: возьмем все возможные тексты размера 8 бит, сгенерируем произвольный ключ длины 128 бит, зафиксируем произвольной константой первую часть настройки и будем проведить шифрование алгоритмом FEA-1 всего в два раунда. Зафиксируем число $\alpha \in \overline{1{,}15}$ и зададим две маски: $\alpha|0$ и $0|\alpha$. Для каждой второй части настройки сгенерированной произвольно 1000 раз, для каждого открытого текста из множества производим зашифрование, с помощью полученного шифртекста получаем квадрат оценки $\hat{\hat{\delta}}$ для первой маски (и на входе, и на выходе маска берется одна и та же) и квадрат оценки для второй маски, усредняем оценки по всем настройкам.

$$1/\mathbb{E}(\delta^2) = \sqrt{N}^{r/2} = (2^4)^{2/2} = 2^4, \mathbb{E}(\delta^2) = 0.0625$$

Реализацию эксперимента можно найти по ссылке. Для каждого α результаты следующие:

α	$\delta(0\|\alpha)$	$\delta(\alpha\|0)$
1	0.0625	0.0863125
2	0.0625	0.0571562
3	0.765625	0.330266
4	0.25	0.0562344
5	0.5625	0.334469
6	0.765625	0.322062
7	1	0.6155
8	0.015625	0.0683437
9	0.5625	0.322297
10	0.25	0.328

3.1 Выводы

В результате работы был разобран линейный метод криптографического анализа шифров, сохраняющих формат. Для проверки теоритических изысканий была создана программа для проведения эксперимента по нахождению линейного соотношения между открытыми текстами, к которым применятся алгоритм шифрования FEA-1, и соответствующими шифртекстами. Результаты эксперимента коррелируют с теорией, однако требуется дальнейшее изучение с целью обобщения теоритических результатов и описания способов атак на FPE шифры.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 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).

2. Алексеев Е.К., Ахметзянова Л.Р., Елистратов А.А., Никифорова Л.О. Шифрование, сохраняющее формат: задачи, подходы, схемы. — 2021. — Режим доступа: https://www.ruscrypto.ru/resource/archive/rc2021/files/02_alekseyev_akhmetzyanova_elistratov_nikiforova.pdf (дата обращения: 25.05.2022).

3. (NIST) Morris Dworkin. Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption. — 2016. — Access mode: https://csrc.nist.gov/publications/detail/sp/800-38g/final (online; accessed: 25.05.2022).

4. Jung-Keun Lee, Bonwook Koo, Dongyoung Roh et al. Format-Preserving Encryption Algorithms Using Families of Tweakable Blockciphers, Ed. by Jooyoung Lee, Jongsung Kim. — Cham : Springer International Publishing, 2015.

5. 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.