

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

головой 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.....	4
1 Шифрование с сохранением формата.....	5
1.1 Описание концепции.....	5
1.2 Действующие стандарты.....	5
1.2.1 FEA-1.....	6
2 Линейный метод.....	8
2.1 Теоритические выкладки из [1].....	9
2.1.1 Теорема.....	10
2.1.2 Применение.....	10
3 Эксперимент.....	12
3.1 Выводы.....	12
Список использованных источников.....	14

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 хранится 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 сохранением формата (format-preserving encryption, FPE), так как в отличие от традиционных механизмов шифрования, оно, во-первых, позволяет программам, обрабатывающим данные как переменные заданного типа, так же успешно обрабатывать и зашифрованные данные, и, во-вторых, позволяет скрыть сам факт шифрования. В 2021 году Тим Бейн, аспирант Лёвенского 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]

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

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 (NSA), 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), M - блок умножения на заданную матрицу.

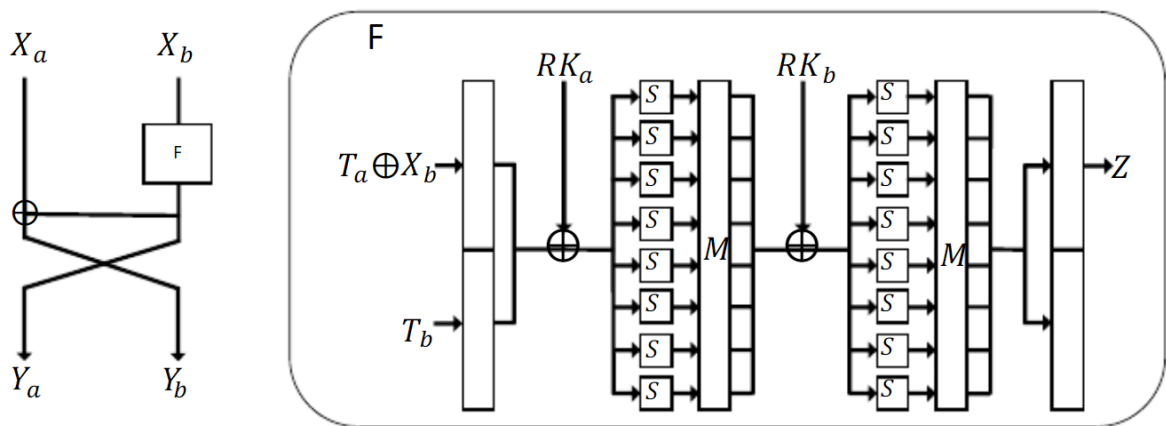


Рисунок 1 — Структура итерации FEA, на основе сети Фейстеля

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

Линейный метод криптографического анализа состоит из двух этапов:

а) Нахождение линейного статистического аналога для части исходного блочного шифра. Это линейное соотношение связывает входные и выходные значения выбранной части алгоритма. Оно должно выполняться с вероятностью заметно отличающейся от случайной для возможности отличия этих двух вариантов событий.

б) Отбрасывание ложных ключей с использованием найденного вероятностного соотношения.

Перейдем к описанию метода:

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

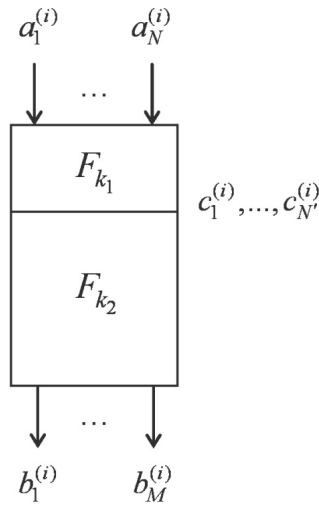


Рисунок 2 — Схема разбиения алгоритма на два блока для проведения линейного криптографического анализа

них используется часть исходного ключа K_1 , во второй, соответственно, K_2 (при этом K_1 может частично совпадать с K_2). $c_1^{(i)}, \dots, c_{N'}^{(i)}$ - промежуточный шифртекст, между двумя блоками шифропреобразования; $b_1^{(i)}, \dots, b_N^{(i)}$ - известный итоговый шифртекст, $a_1^{(i)}, \dots, a_N^{(i)}$ - известный открытый текст, $i \in \overline{1, T}$, где T - количество материала. Пусть также найдено линейное соотношение:

$$c_1^{(i)}L'_1 + \dots + c_N^{(i)}L'_N \simeq b_1^{(i)}L''_1 + \dots + b_N^{(i)}L''_N, \quad (1)$$

которое, **независимо от значения** K_2 , выполняется с вероятностью $P = \frac{1+\delta}{2}$, где $\delta \neq 0$. Булевы величины L'_j и L''_s , $j, s \in \overline{1, N}$ — маска найденного линейного соотношения.

Пусть K'_1 - доля ключа K_1 , от которой зависит левая часть в соотношении 1. Если при опробовании K'_1 выполнимость соотношения с вероятностью $P = \frac{1+\delta}{2}$ не подтверждается, то соответствующее значение K'_1 отбраковывается. Чем больше при этом T и $|\delta|$, тем большая доля значений K'_1 будет отбракована, вплоть до однозначного определения K'_1 .

Для проверки выполнения соотношения вычисляется оценка $\hat{\delta} = 2 \cdot \frac{\sum_{i=1}^T \left(c_1^{(i)}L'_1 + \dots + c_N^{(i)}L'_N == b_1^{(i)}L''_1 + \dots + b_N^{(i)}L''_N \right)}{T} - 1$, где функция под знаком суммы принимает значение 1 при равенстве левой и правой частей, что означает выполнение линейного соотношения для данной пары открытого и зашифрованного текстов, и 0 при неравенстве.

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].

2.1 Теоритические выкладки из [1]

Пусть $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ - 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:

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

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 $2^{-32-n_1/2}$.

2.1.1 Теорема

Пусть получена δ для некоторого линейного соотношения равномерно распределенной функции $F : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$. Случайная величина $2^{n-1}(\delta + 1)$ имеет биномиальное распределение с математическим ожиданием 2^{n-1} и дисперсией 2^{n-2} . В частности, при $n \rightarrow \infty$ распределение $2^{n/2}\delta$ сходится к стандартному нормальному распределению $\mathcal{N}(0,1)$.

2.1.2 Применение

Пусть $r \geq 2$ - 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 2.1.1, N — мощность множества текстов. Переменные δ_i будем считать независимыми, что следует из предположения о независимости раундовых функций F_1, F_3, \dots, F_{r-1} .

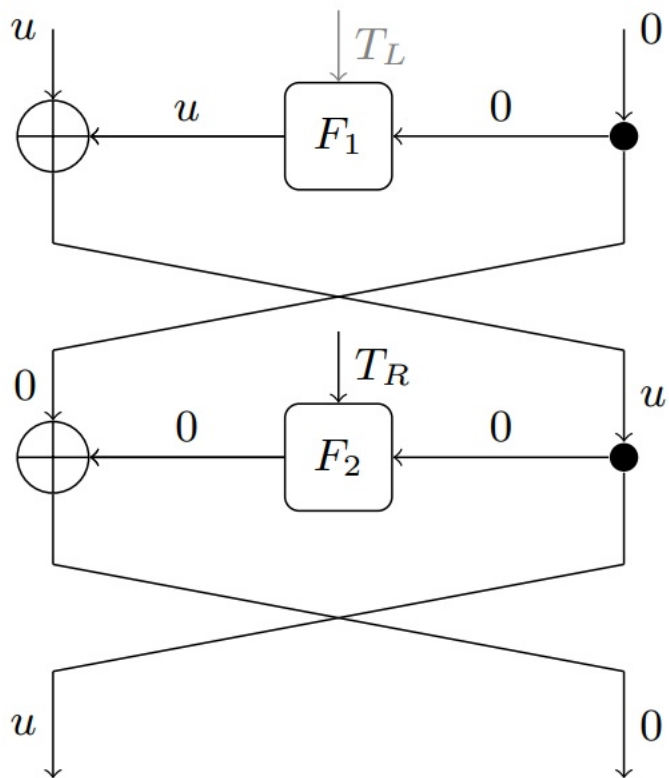


Рисунок 3 — Две итерации FEA-1

Эвристически вычислимо, что для FEA-1

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

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 проводить 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 $\hat{\delta}$ 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 теоретических изысканий была создана программа для проведения эксперимента

по нахождению линейного соотношения между открытыми текстами, к которым применяется 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.