

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение   
высшего образования

**«МИРЭА – Российский технологический университет»**

**РТУ МИРЭА**

Институт искусственного интеллекта   
Базовая кафедра №252 – информационной безопасности

**курсовая работа по предмету**

**«КРИПТОГрафические методы защиты информации»**

**Тема работы:** Криптоанализ алгоритма симметричного шифрования SM4

|  |  |  |
| --- | --- | --- |
| **Студент группы ККСО-01-20** |  | *Семин В.В.* |
| **Руководитель** |  | *Старший преподаватель*  *Хомутов Дмитрий*  *Геннадьевич* |

ОГЛАВЛЕНИЕ

**Элементы оглавления не найдены.**

# 1 ОПИСАНИЕ АЛГОРИТМА ШИФРОВАНИЯ

Описание алгоритма представлено в стандарте [1]. Алгоритм имеет размер блока 128 бит и размер ключа 128 бит. Его структура – несбалансированная сеть Фейстеля (за каждый раунд преобразуется 32 бита блока) с 32 раундами.

## 1.1 РАУНДОВАЯ ФУНКЦИЯ

Раундовая функция имеет вид (1.1).

|  |  |
| --- | --- |
|  | (1.1) |

Здесь – блок открытого текста, – раундовый ключ, – подстановка.

Подстановка является суперпозицией подстановки линейной подстановки и нелинейной подстановки , то есть .

В свою очередь имеет вид (1.2), а имеет вид (1.3).

|  |  |
| --- | --- |
|  | (1.2) |
|  | (1.3) |
| Результатом является блок зашифрованного текста |  |

## 1.2 РАСШИРЕНИЕ КЛЮЧА

Расширение ключа происходит по формуле (1.4)

|  |  |
| --- | --- |
|  | (1.4) |

Здесь , где – ключ, – константы; – константы; - подстановка, идентичная подстановке за исключением того, что вместо используется , имеющая формулу (1.5).

|  |  |
| --- | --- |
|  | (1.5) |

# 2 ОПИСАНИЕ АЛГОРИТМОВ АТАКИ

## 2.1 ЛИНЕЙНАЯ АТАКА НА 22-РАУДОВЫЙ SM4

Статья [2] предлагает линейную атаку на 22 раундовый SM4. Для этого сначала строится линейная аппроксимация раундов 2-19, описанная уравнениями (2.1) – (2.6).

|  |  |
| --- | --- |
| ; раунд 5 | (2.1) |
| ; раунд 6 | (2.2) |
| ; раунд 10 | (2.3) |
| ; раунд 11 | (2.4) |
| ; раунд 15 | (2.5) |
| ; раунд 16 | (2.6) |

Здесь С учётом того, что равно :

|  |  |
| --- | --- |
| ; раунд 5 | (2.7) |
| ; раунд 6 | (2.8) |
| ; раунд 10 | (2.9) |
| ; раунд 11 | (2.10) |
| ; раунд 15 | (2.11) |
| ; раунд 16 | (2.12) |

Сложив уравнения (2.7) – (2.12), получим уравнение линейной аппроксимации раундов 2-19 (2.13).

|  |  |
| --- | --- |
|  | (2.13) |

Автор утверждает, что данная аппроксимация имеет отклонение .

Далее левая часть преобразуется для расширения на 22 раунда:

|  |  |
| --- | --- |
|  | (2.14) |

Здесь – правая часть уравнения (2.13), – результат шифрования 22 раундовым алгоритмом.

Выражение (2.14) имеет форму (2.15)

|  |  |
| --- | --- |
|  | (2.15) |

– функция чётности выражения, зависящая от открытого текста и закрытого. – функция, являющаяся композицией нескольких функций .

Так как в данной атаке количество активных s-боксов 6 для раундов 1, 20 и 8 для 21,22, то зависит от 112 битного вектора , где , , , . Соответственно, представляется формулой (2.16).

|  |  |
| --- | --- |
|  | (2.16) |

Алгоритм атаки следующий:

1. Инициализировать вектор из элементов, каждый из которых соответствует возможному значению битов части текста/шифртекста, использующегося в вычислении .
2. Для каждой пары векторов , вычислить чётность (чётность равна 1, если равно 0, -1 иначе). Сохранить результат в матрице , то есть
3. Вычислить чётность для каждого и прибавить её к соответствующему элементу .
4. Вычислить отклонение .
5. Если , то выбрать соответствующий вариант ключа и предположить

Если , то выбрать соответствующий вариант ключа и предположить

1. После получения значений остаётся перебрать 15 бит для получения ключа.

## 2.2 ДИФФЕРЕНЦИАЛЬНАЯ АТАКА НА 23-РАУНДОВЫЙ SM4

В работе [5] показана дифференциальная атака на 23-раундовый SM4, которая строится на основе 19-раундовой дифференциальной характеристики, рассчитанной авторами статьи.

Данная характеристика выглядит следующим образом. Пусть . Определим ( – это вероятность того, что при разности входов функцией будет получена разность выходов ). Автор статьи утверждает, что множество имеет мощность . Это означает, что возможно получить примерно характеристик данным способом. Определим . Таблица в Приложении А показывает данные характеристики более подробно. В ней , – входная разность и – выходная разность.

Атака на 23-раундовый SM4 начинается с применения дифференциальной характеристики к первым 19 раундам. Для этого понадобится выбрать открытые текста с разностью , где . Сгенерируем два набора и , в каждом из которых открытых текста (формула 2.17).

|  |  |
| --- | --- |
|  | (2.17) |

Здесь – 3 фиксированных слова, – принимают все возможных значений. Назовём пару структурой. Одна структура содержит различных открытых текстов и может создать около кандидатов-пар открытых текстов.

После применения характеристики выходная разность раунда 18 – это . А значит, входная разность функции в раунде 19 равна . Для S-бокса SM4 существует 127 возможных выходных разностей для каждой ненулевой входной разности. Следовательно, функция имеет примерно возможных вариантов. Пусть – множество всех вариантов. Тогда выходные разности последующих раундов (19-22) должны принадлежать соответственно, где обозначает неизвестное слово.

Алгоритм атаки:

1. Выбрать структур. Мы получим пар открытых текстов, используя открытых текстов.
2. Для каждой пары открытых текстов вычислить разность соответствующих шифртекстов и проверить, принадлежит ли первое слово шифртекста множеству . Если это не так, то отбрасываем пару. После этого должно остаться примерно пар.
3. Для каждого предположения о значении нулевого байта , то есть , делаем следующее:
   1. Для каждой оставшейся пары шифртекста пусть , . Частично расшифруем по формуле (2.18), чтобы получить разность нулевого S-бокса в раунде 22., и вычислим по формуле (2.19). Если предположение о значении верно, то для правильных пар . Если это не так, отбрасываем пару. После этого теста должно остаться примерно пар шифртекстов.

|  |  |
| --- | --- |
|  | (2.18) |
|  | (2.19) |

* 1. Для каждого из оставшихся 3 байт , 4 байт , 4 байт продолжаем делать предположения. Для каждого из этих предположений также вычисляем и . Как и в пункте а отбрасываем, неподошедшие пары на каждом шаге. После этих тестов должно остаться примерно пар шифртектов.
  2. Предположим значение . Поступим так же, как в пункте а. Примерно часть пар уже была отброшена, из-за выбора множества . Следовательно, после теста около пар шифртестов останется.
  3. Для каждого из оставшихся 3 байт делаем предположения и повторяем аналогично шагу а. После этого должно остаться около пар шифртекстов.

1. Выведите 128-битное предположение о значении , если оно имеет максимальное число оставшихся пар после шага 3.

# 3 АНАЛИЗ СЛОЖНОСТИ РАССМОТРЕННЫХ АТАК

## 3.1 АНАЛИЗ СЛОЖНОСТИ ЛИНЕЙНОЙ АТАКА НА 22 РАУДОВЫЙ SM4

Введём несколько определений из статьи [3].

Циркулятная матрица. Матрица является циркулянтной, если каждый вектор строки (столбца) сдвинут на один элемент вправо относительно предыдущего вектора строки (столбца).

Блочно-циркулянтная матрица. Матрица является m-блочно-циркулянтной, если она является циркулянтной по блокам, и количество блоков в каждой строке (или столбце) равно m.

Уровневая циркулянтная матрица.

1. Матрица является циркулянтной уровня 1 типа (n), если она является циркулянтной размером n.
2. Матрица является циркулянтной уровня 2 типа (m, n), если она является m-блочно-циркулянтной, и каждый блок сам является циркулянтной матрицей размером n.
3. Матрица является циркулянтной уровня 3 типа (m, n, o), если она является m-блочно-циркулянтной, блоки которой являются циркулянтными уровня 2 типа (n, o).
4. Матрица является циркулянтной уровня q типа (m, n, o, ...), если она является m-блочно-циркулянтной, блоки которой являются циркулянтными уровня q-1 типа (n, o, ...).

Рассмотрим потребление памяти и времени для каждого пункта алгоритма атаки, описанное в [2].

1. Для того, чтобы хранить вектор , состоящий из элементов требуется бит памяти.
2. Матрица M имеет циркулянтную природу. Это следует из утверждения, рассмотренного в [3], о том, что матрица размерами , определённая как , где некоторая функция, является циркулянтной уровня типа . Следовательно, расчёт и хранение всей таблицы не требуется. С учётом этого необходимо вычислить различных значений . Исходя из формулы (2.16), на это понадобится операций зашифрования 22-раундовым SM4, что примерно равно операций. Кроме этого, нужно выделить бит для хранения полученного результата.
3. В шаге 3 используются простые операции извлечения битов и инкрементации. На них понадобится операций.
4. В шаге 4 при прямом подходе необходимо операций. Однако, циркулянтная природа позволяет значительно уменьшить количество операций. Для этого можно использовать методы, описанные в [3]. При этом сложность шага 4 уменьшается до
5. Шаг 6 имеет сложность

В результате, для атаки требуется примерно операций зашифрования 22-раундовым SM4, арифметических операций, бит памяти. Согласно лемме, рассмотренной в [4], понадобится пар открытый/закрытый текст для высокой вероятности успеха .

## 3.2 АНАЛИЗ СЛОЖНОСТИ ДИФФЕРЕНЦИАЛЬНОЙ АТАКИ НА 23-РАУНДОВЫЙ SM4

Рассмотрим анализ сложности алгоритма из [5].

Установлено, что для большинства дифференциальных характеристик, описанных в Приложении А, вероятность . Для остальных она больше, чем . Так что с точки зрения атакующего можно предположить, что вероятность каждой дифференциальной характеристики .

Предполагается, что для корректного ключа останется примерно правильных пар для корректного ключа. Однако, для неправильного предположения о подключах ожидаемое количество оставшихся пар после шага 3 алгоритма примерно .

Для данного анализа будет использована концепция «отношение сигнал-шум». Для определения подходящего значения . Отношение сигнал-шум определено, как отношение вероятности правильного ключа быть предложенным правильной парой к вероятности случайного ключа быть предложенным случайной парой при заданной начальной разности. Отношение сигнал-шум может быть вычислено по формуле (3.1).

|  |  |
| --- | --- |
|  | (3.1) |

Здесь – количество предположенных битов ключа, – вероятность дифференциальной характеристики, – среднее число ключей, предложенное учтённой парой и – отношения количества учтённых пар ко всем парам.

В рассматриваемой атаке предполагается значение 128 бит подключа. Вероятность каждой дифференциальной характеристики . Для каждого теста в шаге 3 есть возможных ключей, и учтённая пара должна удовлетворять 32-битному условию., поэтому . В шаге 2 11-битное условие используется для того, чтобы отбросить пары, поэтому . Следовательно отношение сигнал-шум равно

При этом значении примерно 3-4 правильных пар понадобится для успешной дифференциальной атаки.

Поэтому, выбираем . При этом значении ожидается примерно оставшихся текстов при неправильном ключе и 4 оставшихся текста при правильном.

Следовательно, для успешной атаки понадобится выбранных открытых текстов.

Основную часть временной сложности составляет шаг 3. В пункте 3(а) пар шифртекстов обрабатываются с кандидатов на . Так что сложность примерно 23-раундовых шифрования SM4. Аналогично, сложность 3(б) примерно , 3(в) , 3(г) . Следовательно, общая временная сложность около 23-раундовых SM4 зашифрований.

# 4 ПРАКТИЧЕСКАЯ ЧАСТЬ

## 4.1 ПРОВЕРКА ЛИНЕЙНОЙ АТАКА НА 22 РАУДОВЫЙ SM4

Проверка данного алгоритма была начата с проверки правдивости перехода от уравнения (2.13) к уравнению (2.14). Действительно, переход оказался верным:

Кроме этого, проведена проверка верности выражения (2.14) с помощью программной реализации алгоритма шифрования SM4 на языке программирования Python 3.12. Для этого собрано 10 выборок по 10000 случайных пар текст/шифртекст, зашифрованных на одном ключе, и для каждой выборки подсчитана доля совпадений левой и правой части уравнения (2.14). Результаты представлены в Таблице 4.1

Таблица 4.1 – Результаты проверки верности уравнения (2.14)

|  |  |
| --- | --- |
| Выборка | Доля совпадений |
| 1 | 0,41050 |
| 2 | 0,41004 |
| 3 | 0,40998 |
| 4 | 0,40827 |
| 5 | 0,40914 |
| 6 | 0,40767 |
| 7 | 0,40970 |
| 8 | 0,40966 |
| 9 | 0,41024 |
| 10 | 0,40886 |

Доля совпадений во всех выборках примерно одна и та же. Это показывает, что линейная аппроксимация, представленная уравнением (2.14), действительно приближает 22-раундовый SM4 с высокой точностью.

.

# СПИСОК ЛИТЕРАТУРЫ

1. GMT 0002-2012 SM4分组密码算法 [Электронный ресурс]. Режим доступа: <https://github.com/guanzhi/GM-Standards/blob/master/GMT%E5%AF%86%E7%A0%81%E8%A1%8C%E6%A0%87/GMT%200002-2012%20SM4%E5%88%86%E7%BB%84%E5%AF%86%E7%A0%81%E7%AE%97%E6%B3%95.pdf> (Дата доступа 14.03.24).
2. Taehyun Kim, Jongsung Kim, Seokhie Hong, Jaechul Sung. Linear and Differential Cryptanalysis of Reduced SMS4 Block Cipher [Электронный ресурс]. Режим доступа: https://eprint.iacr.org/2008/281.pdf (Дата доступа 14.03.24).
3. Collard, Baudoin; Standaert, François-Xavier; Quisquater, Jean-Jacques. Improving the Time Complexity of Matsui's Linear Cryptanalysis [Электронный ресурс]. Режим доступа: <http://hdl.handle.net/2078.1/81795> (Дата доступа 24.05.24)
4. M. Matsui, Linear Cryptanalysis Method for DES Cipher [Электронный ресурс]. Режим доступа: <https://link.springer.com/chapter/10.1007/3-540-48285-7_33> (Дата доступа 25.05.24)
5. Bozhan Su, Wenling Wu, Wentao Zhang. Differential Cryptanalysis of SMS4 Block Cipher [Электронный ресурс]. Режим доступа: <https://eprint.iacr.org/2010/062.pdf> (Дата доступа 25.05.24)

# ПРИЛОЖЕНИЕ А

Таблица А.1 – Семейство 19-раундовых дифференциальных характеристик.

| Раунд |  |  |  |  | Вероятность |
| --- | --- | --- | --- | --- | --- |
| 0 |  |  |  |  |  |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  | 1 |
| 3 |  |  |  |  | 1 |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  | 1 |
| 7 |  |  |  |  | 1 |
| 8 |  |  |  |  |  |
| 9 |  |  |  |  |  |
| 10 |  |  |  |  | 1 |
| 11 |  |  |  |  | 1 |
| 12 |  |  |  |  |  |
| 13 |  |  |  |  |  |
| 14 |  |  |  |  | 1 |
| 15 |  |  |  |  | 1 |
| 16 |  |  |  |  |  |
| 17 |  |  |  |  |  |
| 18 |  |  |  |  | 1 |
| 19 |  |  |  |  | 1 |