**Лекция 12.** Линейный криптоанализ. Построение линейных уравнений.

**Линейный криптоанализ.**

Долгое время криптография оставалась секретной наукой, в тайны которой был посвящен лишь узкий круг лиц. Это было естественно. Так как в первую очередь она была направлена на сохранение государственных секретов. Ситуация стала меняться во второй половине ХХ века с появлением персональных компьютеров. Когда практически каждый человек получил возможность оперировать электронной информацией, возникла естественная потребность как-то защищать эту информацию от посторонних глаз. На сегодняшний день наука криптография развивается очень стремительно. Связано это с тем, что в последние годы данная область знаний стала открытой. Если раньше созданием и анализом шифров занимались лишь секретные государственные структуры, то в наши дни любой желающий может беспрепятственно овладеть азами данной науки. Кроме того, быстрое развитие современных информационных технологий также делает криптографию востребованной. Как следствие появляются все новые и новые шифры, предлагаемые авторами разных стран, направленные на усиление секретности данных, шифруемых с их помощью.

Широкое распространение получило использование симметричной криптографии, а несколько позднее и ассиметричной. В 1976 году в США был утвержден стандарт шифрования данных DES (Data Encryption Standard), который использовался довольно длительное время (более 20 лет). Естественно, что у людей возникло желание проверить: а действительно ли предлагаемые алгоритмы для шифрования конфиденциальных данных обеспечивают сохранность информации? Для того, чтобы ответить на этот вопрос необходимо было провести ряд достаточно сложных исследований. Так, исследования в области анализа стойкости шифров постепенно стали причиной того, что в криптологии выделилось два родственных направления, теснейшим образом связанных между собой: криптография и криптоанализ. Прослеживая историю развития этих направлений, можно сказать, что одним из блочных алгоритмов наиболее часто подвергавшийся различного рода атакам является алгоритм шифрования DES. Именно для анализа этого алгоритма шифрования были разработаны такие мощные атаки как линейный и дифференциальный криптоанализ, которые в дальнейшем стали применяться к целому классу блочных шифров. В настоящий момент, знание азов применения данных двух методов анализа, позволяет еще на этапе проектирования шифров заложить в них избыточную устойчивость и заведомо пресечь возможность применения данных методов анализа к вскрытию зашифрованной информации. В настоящей работе предлагается рассмотреть основные принципы, лежащие в основе метода линейного криптоанализа и оценить степень его эффективности.

**1. Основные сведения о линейном криптоанализе**

Метод линейного криптоанализа впервые был предложен в начале 90-х годов XX века японским ученым М. Матсуи (Matsui). В своей работе М. Матсуи показал, как можно осуществить атаку на алгоритм шифрования DES, сократив сложность анализа до 247. Существенным недостатком метода стала необходимость иметь в наличии большой объем данных, зашифрованных на одном и том же секретном ключе, что делало метод малопригодным для практического применения к вскрытию шифра. Однако, если предположить, что к аналитику в руки попал шифрованный текст, содержащий важные сведения, а также некий черный ящик (устройство или программа), который позволяет выполнить любое число текстов, зашифрованных с помощью известного алгоритма шифрования на секретном ключе, не раскрывая при этом самого ключа, то применение метода линейного криптоанализа становится вполне реальным. Позднее М. Матсуи усовершенствовал свою атаку и показал, как можно понизить сложность анализа до 243. Для алгоритма DES метод линейного анализа остался скорее задачей гипотетической, ввиду того, что для анализа требуется огромный объем информации, зашифрованной на одном и том же секретном ключе. Перехватить такой объем информации в реальных условиях практически невозможно. Знание механизмов работы метода линейного криптоанализа позволяет криптографам еще на этапе проектирования крптоалгоритмов обеспечить стойкость шифров. Вот почему так важно уметь применять известные методы криптоанализа на практике.

**2. Определение линейного аналога и его эффективности**

Итак, рассмотрим основные понятия, связанные с методом линейного криптоанализа. Любой алгоритм шифрования в самом общем виде можно представить как некоторую функцию E (от англ. Encryption – шифрование), зависящую от входного сообщения Х, секретного ключа К и возвращающую шифрованное сообщение Y:

Y = E(X, K). (1)

Зная само преобразование Е и входное сообщение Х, нельзя однозначно сказать каким будет выходное сообщение Y. В данном случае нелинейность функции (1) зависит от внутренних механизмов преобразования Е и секретного ключа К.М. Матсуи показал, что существует возможность представить функцию шифрования (1) в виде системы уравнений, которые выполняются с некоторой вероятностью р. При этом для успешного проведения анализа вероятность уравнений р должна быть как можно дальше удалена от значения 0,5 (то есть приближаться либо к 0 либо к единице). Так как уравнения, получаемые в ходе анализа криптоалгоритма, являются вероятностными, то их стали называть линейными статистическими аналогами.

**Определение 1**. Линейным статистическим аналогом нелинейной функции шифрования (1) является величина Q, равная сумме по модулю два скалярных произведений входного вектора Х, выходного вектора Y и вектора секретного ключа К соответственно с двоичными векторами a, β и g, имеющими хотя бы одну координату равную единице:

isch01.wmf,

в том случае, если вероятность того, что Q = 0 отлична от 0,5 (Р(Q = 0) ≠ 0,5).

В отличие от дифференциального криптоанализа, в котором большое значение вероятности гарантирует успех атаки, в линейном криптоанализе успех анализа может быть обеспечен как уравнениями с очень большой вероятностью, так и уравнениями с очень маленькой вероятностью. Для того, чтобы понять, какое из возможных уравнений лучше всего использовать для анализа используют понятие отклонения.

**Определение 2**. Отклонением линейного статистического аналога называют величину h = |1 – 2 р|, где р – вероятность, с которой выполняется линейный аналог.

Отклонение определяет эффективность линейного статистического аналога. Чем отклонение больше, тем выше вероятность успешного проведения анализа. Фактически отклонение показывает насколько вероятность статистического аналога отдалена от значения р = 0,5.

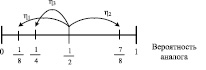
Рассмотрим пример. Пусть имеется три линейных аналога, вероятности для которых соответственно равны isch02.wmf, isch03.wmfи isch04.wmf. Определим какой из этих аналогов лучше всего использовать для анализа. Для этого найдем отклонения аналогов:

isch05.wmf;

isch06.wmf;

isch07.wmf.

Видно, что наибольшее отклонение имеют первый и второй аналоги при том, что значение вероятностей у них различны. Для того, чтобы лучше понять смысл значения отклонения обратимся к рис. 1 из которого наглядно видно, что аналоги с различными вероятностями отклоняются на одинаковое расстояние от точки р = 0,5.



Отклонения линейных статистических аналогов

**3. Основные этапы линейного криптоанализа**

Для успешного применения метода линейного криптоанализа необходимо решить следующие задачи:

1. Найти максимально эффективные (или близкие к ним) статистические линейные аналоги. При нахождении аналогов обратить внимание на то, что в них должно быть задействовано как можно больше битов искомого секретного ключа К.

2. Получить статистические данные: необходимый объем пар текстов (открытый – закрытый текст), зашифрованных с помощью анализируемого алгоритма на одном и том же секретном ключе. При этом можно воспользоваться Парадоксом о Днях рождениях для определения минимального объема данных, необходимого для успешного анализа с вероятностью 0,5.

3. Определить ключ (или некоторые биты ключа) путем анализа статистических данных с помощью линейных аналогов.

Первый шаг анализа заключается в нахождении эффективных статистических аналогов. Для алгоритмов шифрования, в которых все блоки заранее известны, этот шаг можно выполнить единожды, основываясь на анализе линейных свойств всех криптографических элементов шифра. В результате анализа должна быть получена система уравнений, выполняющихся с некоторыми вероятностями. Левая часть уравнений должна содержать в себе сумму битов входного и выходного сообщения, правая часть уравнения – биты секретного ключа. Система уравнений должна быть определенной, то есть содержать все биты исходного секретного ключа. Данный этап не является вычислительно сложным, однако требует больших знаний, логики работы и внимательности. Он может быть автоматизирован. Однако при этом необходимо помнить, что для каждого определенного алгоритма шифрования система линейных аналогов строиться всего один раз и в дальнейшем может быть использована для нахождения разных секретных ключей шифрования, которые используются для шифрования данных с помощью анализируемого шифра. Исключение составляют те блочные шифры, в которых присутствуют нефиксированные элементы. Наглядным примером может служить алгоритм шифрования ГОСТ 28147-89, у которого блоки замены могут быть различны. Как следствие, первый шаг по нахождению статистических аналогов надо будет проделывать снова и снова для каждой новой конфигурации блоков замены.

Если первый шаг анализа является чисто теоретическим и полностью зависит от структуры алгоритма, то второй шаг – является исключительно практической частью, которая заключается в анализе известных пар открытый-закрытый текст с помощью полученной ранее системы статистических аналогов. Для этого используется следующий алгоритм.

**Алгоритм**. Пусть N – число всех открытых текстов и Т – число открытых текстов, для которых левая часть линейного статистического аналога равна 0. Рассмотрим два случая.

1. Если Т > N/2, то в этом случае число открытых текстов, для которых левая часть аналога равна нулю, больше половины, то есть в большинстве случаев в левой части аналога появляется значение, равное нулю, то

а) если вероятность этого линейного статистического аналога р > 1/2, это говорит о том что в большинстве случаев правая и левая части аналога равны, а значит левая часть аналога, содержащая биты ключа, равна 0 (К = 0, если р > 1/2).

б) если вероятность этого линейного статистического аналога р < 1/2, это говорит о том что в большинстве случаев правая и левая части аналога не равны, а значит левая часть аналога, содержащая биты ключа, равна 1 (К = 1, если р < 1/2).

2. Если Т < N/2, то в этом случае число открытых текстов, для которых левая часть аналога равна нулю, меньше половины, то есть в большинстве случаев в левой части а) аналога появляется значение, равное единице, то

если вероятность этого линейного статистического аналога р > 1/2, это говорит о том что в большинстве случаев правая и левая части аналога равны, а значит левая часть аналога, содержащая биты ключа, равна 1 (К = 1, если р > 1/2).

б) если вероятность этого линейного статистического аналога р < 1/2, это говорит о том что в большинстве случаев правая и левая части аналога не равны, а значит левая часть аналога, содержащая биты ключа, равна 0 (К = 0, если р < 1/2).

Успех алгоритма возрастает с ростом N и Δ = |1 – 2p|.

Данный алгоритм будет иметь успех при анализе большого числа текстов N. Следовательно, второй шаг анализа является вычислительно сложным. Поэтому для ускорения времени анализа можно и нужно использовать параллельные вычисления.

В результате работы вышеприведенного алгоритма будет получена определенная (а возможно и переопределенная) система уравнений, отражающая взаимосвязь битов ключа. Третий шаг анализа заключается в решении данной системы, например, методом Гаусса, что позволит получить значения битов секретного ключа шифрования.

**Построение линейных уравнений**

Смысл алгоритма состоит в получении соотношений следующего вида:

https://upload.wikimedia.org/math/6/3/1/631ecef3d7d7c51c6cb3708b24d0fdb1.png

где P_n, C_n, K_n— *n*-ые биты текста, шифртекста и ключа соответственно.

Данные соотношения называются линейными *аппроксимациями*. Вероятность **P** справедливости такого соотношения для произвольно выбранных битов открытого текста, шифртекста и ключа примерно равна 1/2. Такими соотношениями, вероятность которых заметно отличается от 1/2, можно пользоваться для вскрытия алгоритма.

Идея линейного криптоанализа — определить выражения вида (1), которые имеют высокую или низкую вероятность возникновения. Если алгоритм шифрования имеет высокую частоту выполнения уравнения (1), или напротив, уравнение выполняется с низкой частотой, то это свидетельствует о бедной способности рандомизации шифра. Если вероятность выполнения уравнения (1) равна *p*, то вероятность смещения *p* − 1/2. Чем больше величина вероятности смещения |*p* − 1/2|, тем лучше применимость линейного криптоанализа с меньшим количеством открытых текстов, необходимых для атаки.

Есть несколько видов атак линейного криптоанализа. Рассматривается алгоритм Мацуи: изначально для каждого раунда шифра находится линейная аппроксимация с наибольшей вероятностью смещения. Полученные выражения объединяются в общую для всех раундов линейную аппроксимацию, вероятность смещения которой позволяет найти *лемма о набегании знаков*.

Далее рассматривается алгоритм нахождения наилучшей линейной аппроксимации для одного раунда. В блочных шифрах анализ преимущественно концентрируется на S-блоки, так как они являются нелинейной частью шифра. Так как S-блок принимает на входе *m* битов и возвращает *n* битов, от криптоаналитика требуется построить 2*m*+*n* аппроксимаций. Чтобы найти вероятность для одной аппроксимации, на вход S-блоку даются все 2*m* возможных входных значений и идёт подсчёт, сколько раз данная аппроксимация верна для входных и выходных битов. Полученное количество совпадений делится на 2*m*. Линейный криптоанализ имеет одно очень полезное свойство: при определённых условиях (например, когда об открытом тексте известно, что он состоит из символов в кодировке ASCII) можно свести соотношение (1) к уравнению вида:

 C_{j_1} \oplus C_{j_2} \oplus \dots \oplus C_{j_b} = K_{k_1} \oplus K_{k_2} \oplus \ldots \oplus K_{k_c}. 

Здесь отсутствуют биты открытого текста, то есть можно построить атаку на основе только шифртекста. Такая атака может применяться к перехваченному шифртексту и является более практичной.

#### Лемма о набегании знаков

Хотя аппроксимацию с наибольшим отклонением от 1/2 не сложно найти для одного раунда, возникает проблема с вычислением вероятности смещения при экстраполировании метода на полнораундовый шифр. В принципе, это потребовало бы от криптоаналитика просмотреть все возможные комбинации открытых текстов и ключей, что невыполнимо. Решение этой проблемы — сделать ряд предположений и приблизить вероятность, используя лемму о набегании знаков. Пусть мы получили линейные аппроксимации F_i\,(1 \leq i \leq n) для различных раундов, которые равны 0 с вероятностью p_i. Тогда вероятность того, что общая линейная аппроксимация F_1 \oplus F_2 \oplus \dots \oplus F_n равняется нулю, равна

P = \frac{1}{2} + 2^{n-1}\prod_{i=1}^{n} (p_i - \frac{1}{2}).