Криптоанализ шифров:

аффинный шифр, простая замена, шифр Виженера

Е. С. Мойсейчик

Аффинный шифр. Пускай известно, что текст был зашифрован аффинным шифром вида:

$$E(x) = (ax + b)\ mod\ m, \quad D(x) = a^{-1}(x - b)\ mod\ m$$

Пусть также известно, что использовался осмысленный английский текст без пробелов, размер алфавита составляет m = 26 символов.

Тогда, исходя из требований к выбору числа a, знаем, что оно является взаимно простым с модулем m, при этом оно меньше модуля, что значительно сокращает пространство ключей.

Имеем ровно 11 чисел, взаимно простых с 26, которые будем проверять.

Для каждого из выбранных чисел a существует ровно 26 вариантов числа b, итого получаем 11*26 = 286 возможных ключей, что является достаточно малым пространством ключей.

Последовательным перебором в циклах переберём все ключи, расшифруем при помощи них тексты и оценим их принадлежность к английскому языку. Для оценки принадлежности текста будем использовать метод n-грамм. Лучший из полученных текстов и будет открытым текстом сообщения.

Метод N-грамм. В первую очередь определим, что n-грамма – это последовательность из n под ряд идущих букв. В языках существуют разные виды n-грамм.

В конкретной реализации криптоанализа шифров будем использовать квадграммы, т.е. сочетания из букв по 4. Этот метод работает эффективнее, чем метод анализа монограмм, т.к. опирается на особенности языка. Использование большего количества символов в n-грамме не оправдано с точки зрения трудоёмкости.

Метод анализа n-грамм представляет собой частотный анализ встречаемости некоторых n-грамм в тексте, для определения его принадлежность к языку, статистика n-грамм которого считается известной.

Для использования метода понадобится словарь квадграмм английского языка, который можно найти в общем доступе. Будем последовательно

проходить по тексту, выделяя группы по 4 символа и проверять, есть ли такой ключ в словаре квадграмм. Если ключ найден, будем добавлять к рейтингу текста рейтинг этой квадграммы.

Таким образом, в результате работы алгоритма будет собрана статистика квадграмм для каждого из предоставленных претендентов на открытый текст. Лучший из отобранных будет представлять собой текст, наиболее близкий к естественному языку.

Подобным образом для отбора можно также использовать словари, содержащие слова языка, однако в данный работе это считается неоправданным с точки зрения трудоёмкости.

Шифр простой замены. Для анализа шифра простой замены также предположим, что используется осмысленная английская речь с алфавитом в 26 символов.

Тогда из механизма работы самого шифра простой замены, представляющего собой перестановку на множестве алфавита, очевидно, что пространство ключей будет равным 26! .

Т.к. пространство ключей достаточно большое, использовать полный перебор представляется слишком долгим, поэтому для обеспечения сокращения затрачиваемого времени, используем вероятностный метод восхождения к вершине.

Метод восхождения к вершине. Данный метод представляет собой алгоритм, основанный на генерации случайного ключа и улучшении его результатов перестановками по 2 элемента.

Ход алгоритма:

1. Исходный алфавит перемешивается в случайном порядке и выбирается в качестве родительского ключа. Текст расшифровывается этим ключом, оценивается его принадлежность к английскому языку.

2. Родительский ключ подвергается небольшим изменениям – в нашем случае перестановке 2 случайных элементов местами. Назовём полученный ключ дочерним. Текст расшифровывается дочерним ключом и оценивается его принадлежность к английскому языку.

3. Если оценка дочернего ключа выше, чем оценка родительского, то он становится родительским и алгоритм повторяется.

Для реализации надёжной расшифровки шифра простой замены просчитаем в шифр-тексте частоты букв и вынесем предположение о соответствии буквам английского языка по частотам. Расшифруем текст и дадим ему оценку. Данную последовательность примем за родительский ключ.

Далее попробуем найти родительский ключ, лучший предыдущего. Для этого попробуем генерировать случайные перестановки на множестве алфавита и расшифровывать текст. Если найдена перестановка, дающая лучший результат, то примем её в качестве родительского ключа и попытаемся улучшить перестановками по 2 случайных символа.

Для обеспечения надёжности анализа, в текущей работе процесс анализа запускается 7 раз, в них используются 100 попыток найти родительский ключ перемешиванием и поиск улучшений дочернего ключа до тех пор, пока 1000 перестановок под ряд не перестанут улучшать ключ.

Итоговый результат выбирается как лучший по оценке квадграммами среди 7 запусков.

Шифр Виженера. Шифр Виженера представляет собой шифрование с помощью m алфавитов, полученных последовательными сдвигами. Т.к. в данном шифре используется некоторый ключ для выбора алфавита шифрования для конкретного символа, для начала попробуем узнать длину этого ключа. Если будет известна длина ключа k, то также будет известен период символов в тексте, которые зашифрованы одним алфавитов по шифру Цезаря, и криптоанализ всего текста будет сводиться к анализу k шифров Цезаря, которые хорошо вскрываются методом частотного анализа монограмм.

Будем последовательно предполагать длину ключа начиная с 2 и производить анализ индекса совпадений, который является методом оценки совпадения распределения частот текста и равномерного распределения.

Опишем алгоритм на k-ом шаге:

1. Предполагаем, что длина ключа равна k. Разбиваем текст на k подстрок следующим образом: в i-ую подстроку входит каждый i-ый символ исходного текста, $i = 0, \dots, k-1$.

2. Для каждой из подстрок оценим индекс совпадений по формуле вида:

$$I.C. = \frac{\sum_{i=A}^{i=Z} f_i(f_i - 1)}{N(N-1)}$$

где f_i – частота i-ой буквы в тексте, N – количество букв в тексте.

3. Индексом текста будем считать средний индекс совпадения по подстрокам.

4. Те длины ключа, для которых будет получено самое большое значения индекса, сохраним как предполагаемые длины и будем проверять.

Далее выберем некоторое количество самых высоко оцененных длин ключа и будем пробовать взламывать полученные шифры Цезаря по одному, что достаточно эффективно можно сделать, применив критерий согласия Пирсона.

Критерий χ^2 Пирсона. Критерий согласия Пирсона показывает, насколько идентичны два распределения. В нашем случае, нас будут интересовать частоты распределения букв в естественном английском языке и частоты распределения букв в предполагаемом открытом тексте.

По своему принципу работы сбор данной статистики похож на частотный анализ монограмм, но позволяет получить немного более точный результат.

Для вычисления значения данной статистики для каждого из полученных шифров Цезаря будем использовать следующую формулу:

$$\chi^2(C, E) = \sum_{i=A}^{i=Z} \frac{(C_i - E_i)^2}{E_i}$$

где C_i – число встречаемости i-ой буквы в тексте, E_i – число ожидаемых встреч этой буквы в тексте такой же длины на английском языке.

В данном случае, чем меньше показатель χ^2 тем более одинаковы сравниваемые распределения, если показатель равен 0, то они полностью идентичны.

При помощи данной статистики будем взламывать полученные шифры Цезаря и собирать расшифрованные подстроки назад в строку. Далее методом

квадграмм оценим полученный результат и выберем лучший, который и будет представлять собой открытый текст.