КРИПТОАНАЛИЗ
ОПТИМИЗАЦИЯ ТЕСТА КАСИСКИ ПОД ПРОГРЕССИВНЫЙ КЛЮЧ

Панкратьев А.С., Болтак С.В.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Болтак С.В. – ассистент

Шифр - один из способов сокрытия информации. Криптоанализ - наука о методах дешифровки информации в случае, когда нет предназначенного для дешифрования ключа. В данной работе разработан алгоритм взлома шифра Виженера с прогрессивным способом генерации ключевой последовательности.

Один из простейших подстановочных шифров в 17 веке предложил Блез Виженер. В качестве ключа шифрования выступает ключевая фраза. Ключевая фраза повторяется до тех пор, пока не достигнет длины шифруемого текста. Каждая буква секретного сообщения получается сдвигом каждой буквы исходного текста на определённое число, задаваемое буквой ключевой фразы (буква А не даёт сдвига, буква Б — сдвиг на одну позицию, В — на две и т.д.). Например, попробуем зашифровать слово «КУПЛЕТ», пользуясь ключевой фразой «АБВ». После повторения ключевой фразы получается следующая ключевая последовательность: «АБВАБВ». Шифрование происходит по следующему алгоритму: Символ «К» не подвергается сдвигу, символ «У» сдвигается на одну позицию, превращаясь в «Ф», символ «П» сдвигается на две позиции, превращаясь в «С». После дальнейшего шифрования, мы в итоге получаем «КФСЛЁФ» [1].

На протяжении нескольких столетий такой способ шифрования считали практически не взламываемым. Однако в 19 веке были предприняты первые попытки взлома данного шифра. Все они были основаны на нахождении длины ключевой фразы, ведь если известна её длина, то зашифрованный текст можно разбить на фрагменты, каждый из которых кодируется обычным шифром Цезаря [1]. В примере выше К,Л – кодируются нулевым сдвигом. У,Е – сдвигом на символ, П,Т – сдвигом на два символа. Если текст достаточно длинный, можно применить частотный анализ и получить исходный текст. Из этого следует, что взлом данного шифра сводится к поиску длины ключевой фразы. Одним из наиболее популярных алгоритмов для этой цели служит тест Касиски [2]. Он применяется для нахождения длины ключа в полиалфавитных шифрах, таких как шифр Виженера. Основная идея алгоритма основана на том, что одинаковые участки открытого текста при шифровании дают одинаковые отрезки криптограмм. Таким образом, найдя расстояние между такими отрезками, мы сможем сделать вывод о длине ключа.

Если для прямого ключа тест работает отлично, то для прогрессивного - классический тест Касиски становится бесполезным. Сама идея использования прогрессивного ключа заключается в циклическом сдвиге символов ключа на одну позицию в упорядоченном алфавите символов при повторном применении ключа. Тогда для ключа «АГЕ» при повторном его использовании по прогрессивной схеме имеем «БДЁ», а при третьем «ВЕЖ», и так далее.

При разработке программного средства для криптоанализа шифра с прогрессивной ключевой последовательностью были рассмотрены различные алгоритмы. Самым очевидным решением является следующее изменение классического теста: при поиске одинаковых *L*-грамм выбирать *L*-граммы, лежащие на расстояниях кратных длине алфавита, так как в этом случае ключ вернется в свое исходное состояние. Недостаток - для текстов на русском языке длина шифротекста должна быть порядка 2000 символов.

В данной работе предлагается следующее решение: для поиска одинаковых *L*-грамм необходимо на каждой итерации сдвигать символы на одну позицию, учитывая номер операции для определения расстояния. Если искомая *L*-грамма начинается с индекса *i* и она повторно найдена в шифротексте, начиная с индекса *j* после *k* сдвигов символов искомой *L*-граммы, то расстояние между *L*-граммами равно остатку от целочисленного деления $(i - j \cdot k)$ на количество букв в алфавите.

Таким образом, алгоритм теста Касиски для прогрессивного ключа будет иметь следующий вид:

1) поиск в тексте повторяющихся *L*-грамм и нахождение расстояний между ними (схема алгоритма приведена на рисунке 1):
   - найти расстояния на *i*-ой операции и занести в список итерационных расстояний;
   - в конце итерации найти наиболее часто встречающееся повторение и занести в список;
   - очистить список итерационных расстояний;
2) среди найденных расстояний поиск наименьшего делителя, данное значение – искомая длина ключевой фразы.

Получив длину ключа, криптоаналитик с помощью частотного анализа может легко получить искомый текст.

Полученный алгоритм отлично справляется с шифротекстами от 300 слов и более с повторяющимися *L*-граммами в исходном тексте и ключом, длиной не больше длины алфавита. Для более маленьких текстов алгоритм так же работает, хоть и с меньшей точностью нахождения ключа.

Схема на рисунке 1 поясняет алгоритм , описанный выше

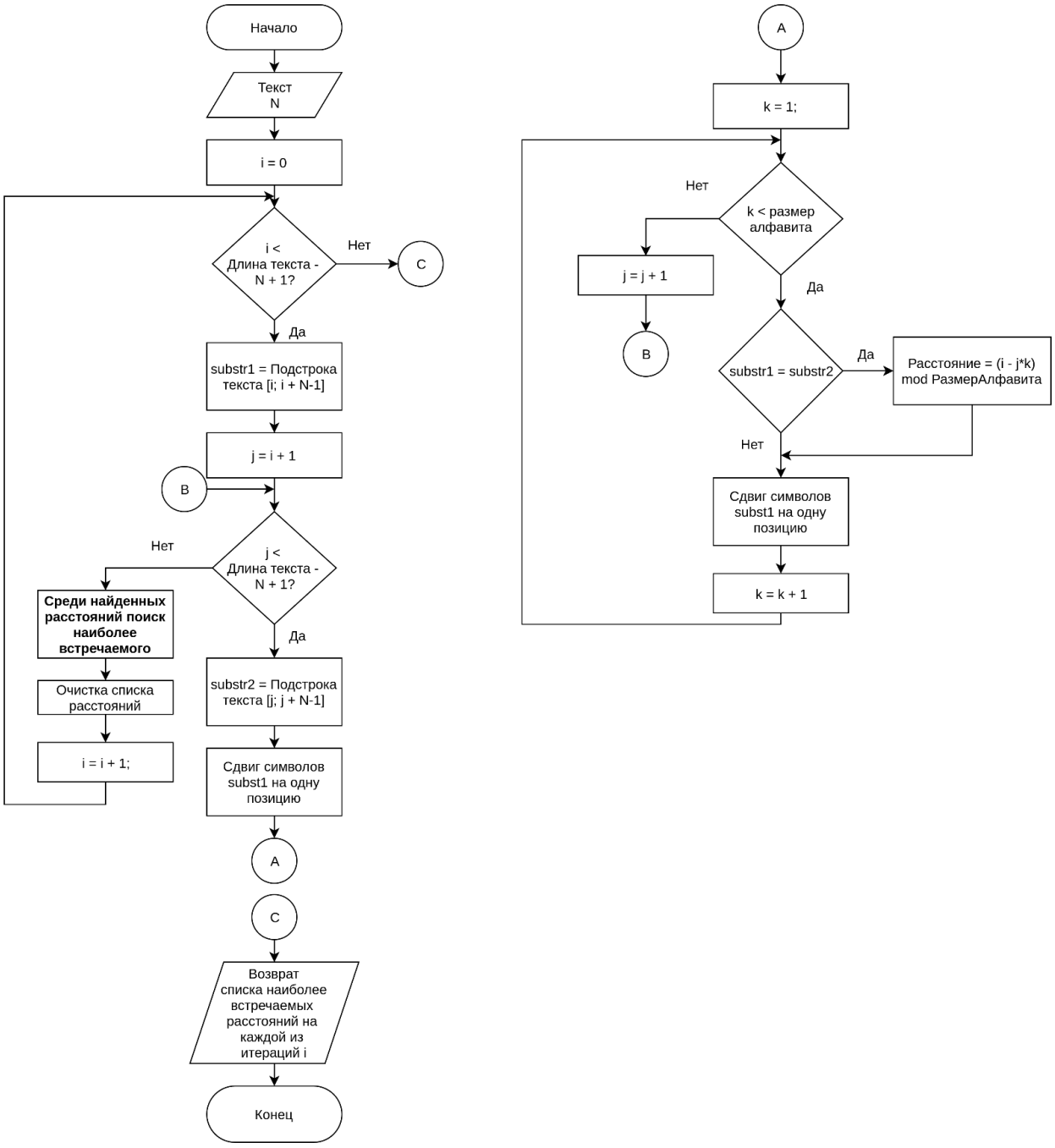

Рисунок 1 – Алгоритм поиска расстояний между *L*-граммами

Тест Касиски позволяет только получить длину ключа, однако прекрасно сочетается с различными методами криптоанализа, предполагающими наличие длины ключевого слова.

Разработанное программное средство позволяет осуществить криптоанализ шифротекста, полученного путём шифрования методом Виженера с прогрессивной ключевой последовательностью.

Список использованных источников:

1. С.Сингх. Тайная история шифров и их расшифровки. — Астрель, 2006. — 447 с.

2. Ярмолик, В. Н. - Элементы теории информации : практикум для студ. спец. «Программное обеспечение информационных технологий» дневн. и дист. форм обуч. / В. Н. Ярмолик, А. П. Занкович, С. С. Портянко. – Минск : БГУИР, 2007. – 39 с.