Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра радиотехнических систем

С.Б. Саломатин

ЗАЩИТА ИНФОРМАЦИИ

МЕТОДИЧЕСКОЕ ПОСОБИЕ

к лабораторной работе
«Криптоанализ алгоритмов защиты информации»
для студентов специальностей
39 01 01 «Радиотехника», 39 01 02 «Радиоэлектронные системы»
дневной формы обучения

Библиотека БГУИР

Минск 2003

УДК 681.3.04 (075.8)
ББК 32. 811 я 73
С 16

Саломатин С.Б.

С 16 Защита информации: Метод. пособие к лаб. работе «Криптоанализ алгоритмов защиты информации» для студ. спец. 39 01 01 «Радиотехника», 39 01 02 «Радиоэлектронные системы» дневной формы обучения / С.Б. Саломатин. – Мн.: БГУИР, 2003. – 20 с.

ISBN 985-444-532-1.

Методическое пособие содержит теоретические сведения о методах криптографического преобразования и дешифрования информации, а также алгоритмы, программы моделирования, реализующие статистические методы криптографического анализа моноалфавитных и многоалфавитных шифров. Приведены содержание и порядок выполнения лабораторной работы.

УДК 681.3.04 (075.8)
ББК 32.811 я 73

© Саломатин С.Б., 2003
© БГУИР, 2003

ISBN 985-444-532-1

Библиотека БГУИР

Содержание

Библиотека БГУИР

Все вещи в мире представляют собой шифр. Вся природа является просто шифром и секретным письмом.
Блез де Виженер «Трактат о шифрах»

1. ЦЕЛИ РАБОТЫ

1. Изучить криптографические методы анализа алгоритмов защиты информации.
2. Исследовать алгоритмы криптоанализа моноалфавитных и многоалфавитных криптосистем.
3. Получить навыки программирования алгоритмов криптоанализа.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

2.1. Моноалфавитные криптографические системы

Моноалфавитные криптосистемы используют простые *операции замены* для шифрования данных.

Шифр Цезаря

Историческим примером шифра замены является шифр Цезаря (I век до н.э.), описанный историком Древнего Рима Светонием. Гай Юлий Цезарь использовал в своей переписке шифр собственного изобретения. Применительно к современному русскому языку он состоял в следующем: выписывался алфавит – А, Б, В, Г,…, затем под ним выписывался тот же алфавит, но с циклическим сдвигом на 3 буквы влево:

АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ
ГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯАБВ.

При зашифровании буква А заменялась буквой Г, Б – Д, В – Е и так далее. Например, слово «ГАЙ» превращалось в слово «УГМ». Получатель сообщения искал эти буквы в нижней строке и по буквам над ними восстанавливал исходное слово. Ключом в шифре Цезаря является величина сдвига нижней строки алфавита. Обобщение шифра Цезаря состоит в использовании произвольного циклического сдвига и различного расположения букв в нижней строке:

$$c_j = i_j + \mathrm{b} \mod N,$$

где c_j – j-й символ шифра; i_j – j-й символ шифруемого сообщения (открытого текста); b – ключ, определяющий величину циклического сдвига. Иногда такой вид шифра называют *аддитивным*.

Дешифрование шифра простой замены

Библиотека БГУИР

Для дешифрования аддитивных шифров замены удобно использовать устойчивые закономерности открытого текста. К наиболее устойчивым закономерностям открытого текста относятся следующие:

- в осмысленных текстах любого естественного языка различные буквы встречаются с разной частотой, при этом относительные частоты букв в различных текстах одного языка близки между собой. То же самое можно сказать и о частотах пар, троек букв открытого текста;
- любой естественный язык обладает избыточностью, что позволяет с большей вероятностью «угадывать» смысл сообщения, даже если часть букв сообщения не известна.

В табл. 1 приведены относительные частоты ν повторения букв алфавита Λ русского языка.

Таблица 1

№ п.п	Λ	ν	№ п.п	Λ	ν	№ п.п	Λ	ν
1	А	0,062	12	Л	0,035	23	Ц	0,004
2	Б	0,014	13	М	0,026	24	Ч	0,012
3	В	0,038	14	Н	0,053	25	Ш	0,006
4	Г	0,013	15	О	0,090	26	Щ	0,003
5	Д	0,025	16	П	0,023	27	Ъ, Ь	0,014
6	Е, Ё	0,072	17	Р	0,040	28	Ы	0,016
7	Ж	0,077	18	С	0,045	29	Э	0,003
8	З	0,016	19	Т	0,053	30	Ю	0,006
9	И	0,062	20	У	0,021	31	Я	0,018
10	Й	0,010	21	Ф	0,002	32	пробел	0,175
11	К	0,28	22	Х	0,009			

Если упорядочить буквы по убыванию вероятностей, то мы получим вариационный ряд:

О,Е,А,И,Н,Т,С,З,В,Л,К,М,Д,П,У,Я,З,Ы,Б,Ь,Г,Ч,Й,Х,Ж,Ю,Ш,Ц,Щ,Э,Ф.

Частотная диаграмма зависит от языка. Например, для английского языка наиболее употребляемые буквы характеризуются следующими относительными частотами (в процентах):

E –12,75; T – 9,25 ; R – 8,50; I – 7,75; H – 7,75; O – 7,50.

Частотная диаграмма является устойчивой характеристикой текста. Из теории вероятностей следует, что при достаточно слабых ограничениях на вероятностные свойства случайного процесса справедлив закон больших чисел, т.е. относительные частоты знаков сходятся по вероятности к значениям их вероятностей. Это верно для последовательности независимых испытаний, для конечной, регулярной однородной цепи Маркова. Эксперименты показывают, что это верно и для открытых текстов.

Библиотека БГУИР

Частотный метод дешифрования шифров простой замены

Метод использует частотные характеристики открытого текста. Если упорядочить по убыванию частоты встречаемости знаков в шифрованном тексте и сравнить с вариационным рядом вероятностей открытого текста, то эти две последовательности будут близки. Скорее всего на первом месте окажется пробел, далее будут следовать буквы О, Е, А, И. В случае, если текст не очень длинный, то необязательно полное совпадение. Но в любом случае в первых и вторых рядах одинаковые буквы будут располагаться недалеко друг от друга, и чем ближе к началу ряда (чем больше вероятность появления знаков), тем меньше будет расстояние между знаками упорядоченной последовательности и вариационного ряда открытого текста.

Аналогичная картина наблюдается и для пар соседних букв, которые называют *биграммами* или *диграфами*. Для получения устойчивой картины длина последовательности должна быть достаточной большой. На сравнительно небольших отрезках анализа картина смазывается. Более устойчивой характеристикой биграммы является отсутствие в осмысленном тексте некоторых биграмм (запретных биграмм, имеющих вероятность, равную практически нулю).

Алгоритм дешифрования шифра простой замены

Алгоритм дешифрования рассмотрим на примере. Пусть имеется следующий шифротекст:

ДОЧАЛЬ ИЬЦИО ЛИОЙО ВНЫИЮШ ХЕМВНЛХЕИ ДОСОЛЬ ЧСО ИА ТЬЖАТСР ЬАС АКЕИОЙО ДОКЩОКЗЖАЙО КПЗ РТАЩ ТПЬЧНАР ТДОТОУН ХЕМВОРНИЕЗ ЕИМОВЛНЯЕЕ РЮУОВ ББЕД СОЙВНМЕЧАТБОГ ТЕТСАЛЮ ЫНРЕТЕС ОС ОТОУ АИИОТСАГ ЕИМОВЛНЯЕЕ АА ЯАИИОТСЕ Е РОЫЛОЦИОТСАГ РПНКАПШЯАР ДО ЫНЖУСА ТРОАГ ЕИМОВНЯЕЕ ДВАЦКА РТАЙО ДОКЧАВБИАЛ УОПШХОА ВНЫИООУ ВНЫЕА РЕКОР ЫНЖЕЖНАЛОГ ЕИМОВЛНЯАА КОБЬЛАИСНПШИНЗ САПАМОТТНЗ САПАРЕЫЕОИИНЗ БОЛДШЭСАВИНЗ БНЦКЮГ РЕК ЕИМОВЛНЯЕЕ УЛААС ТРОЕ ТДАЯЕМЕЧАТБЕА ОТОУАИИОТСЕ Е ФСЕ ОТОЧАИИОТСЕ ТЕПШИО РПЕЗЭС ИН РЮУОВ ЛАСОКОР ХЕМВОРНИЕЗ ЕИМОВЛНЯЕЕ УОПШХОА ЫИНЧАИЕА ЕЛАЭС ОУЪАЛЮ Е СВАУЬАЛНЗ ТБОВОТСШ ДАВАКНЧЕ ХЕМВОРНИИОГ

1. Посчитаем частоты символов в шифрованном тексте.

2. Упорядочим символы по убыванию частот, образуя последовательность:

$$c_1, c_2, c_3, c_4, c_5, \ldots .$$

3. Запишем под образованной последовательностью вариационный ряд вероятностей знаков в открытом тексте.

4. При достаточно большой длине анализируемого текста для того, чтобы из шифротекста получить открытый, достаточно заменить c_1 на первую букву вариационного ряда, c_2 – на вторую букву вариационного ряда и т. д .

Библиотека БГУИР

Такая ситуация будет иметь место для наиболее вероятных букв. Проделав указанные операции, можно обнаружить, что дешифрованный текст не читается. Следовательно, материала для анализа недостаточно. Остается угадывать замену, при этом можно учитывать статистические особенности открытого текста. В шифрованном тексте через пробел скорее всего будет обозначен пробел открытого текста, через букву О скорее всего обозначена О или А; через Е – О, Е, А; через А – Е, А или И и т.д.

5. Выпишем шифротекст, а под ним – колонку наиболее вероятных замен букв. Чем реже встречается буква, тем большей глубины надо анализировать колонку вероятных замен, чтобы была уверенность, что в колонке содержится знак открытого текста.

В рассматриваемом примере колонки взяты глубиной в пять символов, но выписаны только для наиболее частых букв (табл.2). Угаданные буквы для глубины анализа больше 5 символов подчеркнуты

Таблица 2

↓	*Колонки наиболее вероятных замен букв шифротекста*										↓
			о	р			е			е	
			е	в			а			а	
Д	О	Ч	А	Л	Ь		И	Ь	Ц	И	О
	е		и	к			н			н	е
	а		н	м			т			т	а
Дешифрованный текст											
п	о	ч	е	м	у		н	у	ж	н	о

Мультипликативные шифры простой замены

В мультипликативном шифре символ образуется в результате умножения знака открытого текста на ключ

$$c_j = (\text{a}\ i_j) \bmod N,$$

где а – фиксированное число-ключ; N – мощность алфавита.

На первый взгляд, определить ключ шифрования достаточно просто. Действительно, вычислим упорядоченную частотную диаграмму шифротекста и сопоставим ей вариационный ряд вероятностей открытого теста $\{b_j\}$. Тогда, сравнив два наиболее вероятных знаков этих последовательностей, можно оценить ключ шифрования по правилу

$$\hat{a} = c_j b_j^{-1} \bmod N .$$

Однако из теории чисел известно, что обратное число b_j^{-1} может быть вычислено однозначно только когда НОД(b_j, N) = 1. В противном случае возникает неопределенность или многозначность решения. Например, если задача состоит в том, чтобы найти обратное число по модулю 26, то это не однозначная задача для Z[26].

Библиотека БГУИР

Аффинная криптосистема

Криптосистема, которая для шифрования использует преобразование множеств Z/NZ вида

$$C \equiv a\,P + b \bmod N, \qquad (1)$$

где a и b – фиксированные целые числа, образующие ключ шифрования, называется аффинной.

Пример. Зашифруем сообщение “PAYMENQW”, используя 26-значный алфавит английского языка (N = 26). Ключи шифрования a = 7, $в$ = 12.

Оцифровка сообщения дает код : (15. 00. 24. 12. 04. 13. 14. 22.).

Преобразование по правилу (1) приводит к (13. 12. 24. 18. 14. 25. 06. 10.) или в буквенном эквиваленте (NMYSOZGK).

Чтобы дешифровать сообщение, которое зашифровано способом аффинных преобразований, необходимо вычислить

$$P \equiv a` C + b` \bmod N,$$

где $a`$ – обратное к ключу $a$ по mod $N$ и $b` = (-a^{-1})\,b$.

Однозначно обратные числа можно найти, если НОД(a, N) = 1, т.е. a и N являются взаимно простыми числами. В противном случае выразить P через C не удается. Легко убедиться, что в этом случае зашифрованной букве соответствует больше, чем одна буква открытого текста.

Задача криптоанализа

Исследуется гипотеза, что перехваченное шифросообщение получено с помощью аффинного отображения N-значного алфавита. Требуется определить ключ сообщения и прочесть сообщение.

Данную задачу можно решить, используя результаты анализа частоты повторения символов в принятых ранее шифровках.

Пример. Предположим, что при N = 26 (английский алфавит) в ранее принятых шифровках наиболее часто встречается буква “K”, второй по частоте повторения является буква “D”. Логично предположить, что эти буквы соответствуют наиболее часто встречающимся в английском языке буквам “E”, “T”. Расположив буквы по их цифровым эквивалентам и заменив P и C в дешифровальных формулах, получим:

$$10\,a` + b` = 4 \bmod 26,$$

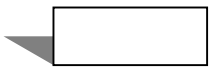
$$2a` + b` = 19 \bmod 26.$$

Вычитая второе равенства из первого, приходим к соотношениям:

Библиотека БГУИР

$$7a` = 11 \bmod 26,$$
$$a` = 7^{-1}\ 11 \bmod 26.$$

Далее, вычисляя $b` = (4 - 10\ a`) \bmod 26$, можно дешифровать сообщение по формуле



$$P = 9C + 18 \bmod 26.$$

В общем случае при криптоанализе аффинных систем знание частотной диаграммы двух наиболее часто встречающихся букв может оказаться недостаточно. Это происходит в том случае, когда несколько возможных преобразований дают один и тот же результат. В этом случае необходимо использовать смысловое дешифрование сообщения.

Диграфные преобразования

Единицы сообщений, составленные из двух букв (блоки), называются диграфами (биграммами). Если весь текст имеет нечетное количество букв, то чтобы получить целое число диграфов, используют добавление в конце текста особой буквы.

Каждому диграфу ставят в соответствие цифровой эквивалент. Простейший способ получения цифрового эквивалента – использование преобразования вида

$$X\ N + Y,$$

где X – цифровой эквивалент первой буквы в диграфе; Y – цифровой эквивалент второй буквы в диграфе.

Например, сообщение диграф “NO” имеет цифровой эквивалент $13\ 26 + 14 = 352$.

Аффинные преобразования диграфных сообщений

Шифрование диграфных сообщений осуществляется по правилу

$$C = a\ P + b \bmod N^2. \qquad (2)$$

Если N не имеет общих множителей с a, b, то возможна однозначная дешифровка сообщения по правилу

$$P = a`\ C + b` \bmod N^2,$$

где $a` = a^{-1} \bmod N^2$ и $b` = -a^{-1}\ b \bmod N^2$.

Далее осуществляется перевод из диграфного блока по правилу

$$C = x`N + y`.$$

После чего определяются буквы по их цифровым эквивалентам x` и y`.

Пример. Если в 26-значном английском алфавите используется шифрование

$$C= 159\ P + 580 \bmod 676,$$

то диграф “ON” имеет цифровой эквивалент $14\ 26 + 13 = 377$ и после зашифровки получаем

$$159\ 377 + 580 \bmod 26^2 \rightarrow 359.$$

Библиотека БГУИР

Криптоанализ диграфных аффинных преобразований

При проведении статистического криптоанализа используется информация о частотном анализе (частоте повторения) диграфов в языке и в перехваченных шифровках. Например, если используется 26-значный английский алфавит, то статистический анализ показывает, что диграфы “TH” ,“HE” – наиболее часто встречающиеся. Знание двух пар диграфов шифровки позволяет (не всегда однозначно) расшировать сообщение.

Пример. В качестве исходных данных используется 27-значный алфавит, в котором 27-м знаком является пробел. Каждый диграф соответствует целому числу в диапазоне (0 – 728), и цифровой эквивалент определяется по формуле (27 x + y). Статистический анализ шифровок показал, что наиболее часто встречаются диграфы вида “ZA”, “IA”, “IW”. Статистический анализ английского языка показывает, что в 27-значном алфавите наибольшую частоту повторения имеют сочетания “E пробел”, “S пробел”, “пробел T”. Предполагается, что используется аффинное отображение для шифрования сообщения.

Задача. Определить ключ дешифрования и прочитать сообщение “NDXBHO”.

Решение. Проверяется гипотеза, что текст и шифрованное сообщение связаны правилом

$$C = a\,P + b \bmod 729,$$

а дешифрование осуществляется по правилу

$$P = a`C + b` \bmod 729,$$

где a, b – ключи шифрования, а a`, b`– ключи дешифрования.

Зная результаты статистического частотного анализа, можно составить три уравнения:

$$675a` + b` = 134 \bmod 729,$$
$$216\,a` + b` = 512 \bmod 729,$$
$$238\,a` + b` = 721 \bmod 729.$$

Решая данную систему уравнений, получим следующие соотношения:

$$459\,a` = 351 \bmod 729,$$
$$437a` = 142 \bmod 729.$$

Первое соотношение приводит к неоднозначному результату (27 решений). Из второго соотношения, используя алгоритм Евклида для определения обратного числа, можно найти

$$a` = 437^{-1}\,142 \bmod 729 \rightarrow 362\,142 \bmod 729 \rightarrow 374;$$
$$b` = 134 - 675\,374 = 647 \bmod 729.$$

Применив ключи дешифрования к диграфам принятого шифрованного сообщения, получим ND = 354→365; XB = 622→724; HO = 203→24. Далее можно записать

$$365 = 13\,27 + 14,$$
$$724 = 26\,27 + 22,$$
$$24 = 0\,27 + 24.$$

Сложив вместе, получим текст сообщения “NOWAY”.

Библиотека БГУИР

Определим ключи шифрования

$$a = a^{`-1} = 374^{-1} = 614 \bmod 729;$$

$$b = -a^{-1}b` = -614\ 647 \bmod 729 = 47 \bmod 729.$$

2.2. Многоалфавитные шифры и методы их анализа

Шифр гаммирования

Пусть буквы алфавита I упорядочены в некотором естественном порядке. Проведем оцифровку букв и слов, поставив в соответствие каждой букве ее цифровой эквивалент (например, порядковый номер буквы в алфавите) и каждому слову – набор соответствующих цифровых эквивалентов букв. Множество всех цифровых слов в алфавите определим как множество конечных последовательностей $(i_1, i_2, ..., i_L)$, $i_j \in L$. Для ключа (гаммы) $g = g_1, g_2, ..., g_L$ определим *шифр гаммирования* как набор чисел

$$f_g(i_1, ..., i_L) = y_1, y_2, ..., y_L,$$

где $y_j = i_j + g_j \mod n$, $n = |I|$.

Шифр Виженера

Шифр Виженера можно считать родоначальником шифров гаммирования. Суть первоначального шифра Виженера можно пояснить следующим образом. Выпишем латинский алфавит:

ABCDEFGHIJKLMNOPQRSTUVWXYZ .

В качестве ключа выберем число 15792. Это число периодически выписывается над буквами открытого текста (одна цифра над буквой). При шифровании буква открытого текста заменяется на букву, стоящую от нее справа (циклически) в алфавите на расстоянии, определяемом соответствующей цифрой ключа. Например, при заданном ключе слово «CIPHER» превращается в шифрованную последовательность: «DNWQGS».

Дальнейшая модернизация привела к шифру модульного гаммирования. Пронумеруем буквы алфавита: A = 01, B = 02, C = 03,…, Z = 26. Проведем оцифровку слова «CIPHER», сопоставив каждой букве слова цифру ее порядкового номера в алфавите:

«CIPHER» → 03.09.16.08.05.18.

Применим операцию модульного сложения по (x + z) mod N, где N – мощность алфавита (для рассматриваемого примера N = 26). Выпишем ключ 15792, периодически повторяя его под оцифрованным открытым текстом, и сложим соответствующие числа. Получим шифротекст 04.14.23.15.07.19., что соответствует сочетанию букв «DNWQGS». При расшифровании ключ вычитается из шифротекста.

Предположим, что алфавит открытого текста $(i_0, i_1, i_2, i_3, ...)$ состоит из n символов. Ключом криптосистемы является последовательность $k_0, k_1, ..., k_L$ из некоторого числа L символов. Каждому знаку открытого текста и ключа поставим в соответствие некоторый вычет по модулю n. Тогда криптосистему Виженера можно получить по правилу

Библиотека БГУИР

$$y(t) = i_t + k_t \mod n .$$

Напомним, что суммирование по модулю часто называют гаммированием, а саму ключевую последовательность – гаммой. Таким образом, систему Виженера можно трактовать как шифр гаммирования, при этом ключ (k_1,…,k_L) рассматривается как гамма (γ_1,…,γ_L).

Дешифрование шифра гаммирования

В основе методов дешифрования шифров гаммирования лежат алгоритмы определения периода гаммы по известному зашифрованному тексту. Для дальнейшего изложения полезно ввести такое понятие, как *индекс совпадения*.

Пусть I – некоторый алфавит. *Индексом совпадения* последовательности $J = i_1, i_2, ..., i_N \in I^N$ называется величина

$$IC(J) = \sum_1^{|I|} \frac{F_i(F_i - 1)}{N(N-1)},$$

где F_i – частота встречаемости буквы i в последовательности J (количество позиций, на которых стоит буква i).

Из определения следует, что индекс совпадения последовательности J равен вероятности $P_J(i_j = i_{j'})$ совпадения букв данной последовательности на случайно и равновероятно выбранных местах $j, j' \in \overline{1, N}, j \neq j'$. Число $F_i(F_i - 1)$ характеризует количество возможных благоприятных событий ($i_j = i_{j'} = i$), а число $N(N\text{-}1)$ – количество всех возможных выборов пар упорядоченных мест (j, j') в последовательности J.

Будем предполагать, что буквы алфавита открытого текста I отождествляются с их номерами (от 0 до $|I|-1$) при расположении букв в стандартном порядке. Через $P_0 = (P_1, ..., P_{|I|})$ обозначим вероятностное распределение на I, где P_i – вероятность буквы i в содержательных открытых текстах. Выборку из вероятностного распределения определим как реализацию случайной выборки $A(N) = (a_1,…,a_N)$ из генеральной совокупности распределения P_0. Для индекса совпадения справедливо следующее соотношение:

$$IC(A(N)) = \sum_1^{|I|} \frac{F_i(F_i - 1)}{N(N-1)} \to \sum_1^{|I|} P_i^2 .$$

Определение периода гаммы по заданному шифротексту

Один из приемом, ускоряющий процесс дешифрования, состоит в определении периода гаммы по заданному шифротексту (либо установлении факта ее непериодичности). Исходными данными для решения этой задачи являются: $A(N) = =(a_1,…,a_N)$ – открытый текст; $G(N) = (\gamma_1,…,\gamma_L)$ – гамма из множества U, $B = b_1,…,b_N$ – зашифрованный текст. Рассмотрим наиболее известные методы оценки периода гаммы.

Библиотека БГУИР

Статистический метод

Статистический метод основан на переборе возможных значений периода и решении для каждого значения периода задачи «*о перекрытии*» – установлении факта того, что два заданных шифротекста получены зашифрованием двух открытых текстов одной гаммой.

Метод чересериодного вычитания

Для проверяемого периода d образуется вспомогательная последовательность

$$b_1 - b_{1+d},\ b_2 - b_{2+d},\ldots,b_j - b_{j+d},\ldots,b_{(l-1)d+r} - b_{ld+r},\quad N = ld + r.$$

Вычитание и сложение проводятся по модулю $|I|$ номеров букв, упорядоченных в естественном расположении. Так как $b_j = a_j + g_j$, то в случае истинного периода d гаммы получаем

$$b_j - b_{j+d} = a_j - a_{j+d}$$

для любого $j \in \{1,\ldots,(l-1)d + r\}$, т.е. вспомогательная последовательность является разностью двух открытых текстов. При случайном выборе этих открытых текстов вспомогательная последовательность тоже является случайной, вероятностное распределение букв (номеров) которой равно P (ГИПОТЕЗА H(0)).

Если d не является истинным периодом гаммы, то

$$b_j - b_{j+d} = a_j + g_j - a_{j+d} - g_{j+d}.$$

Можно допустить, что символы гаммы выбирались при шифровании независимо и вспомогательная последовательность не является разностью двух открытых текстов. В этом случае предполагается, что вспомогательная последовательность является случайной независимой выборкой из равновероятного распределения I (ГИПОТЕЗА (H(1)). Относительно вспомогательной последовательности решается статистическая задача – принятия гипотезы H(0) или H(1). В первом случае величина d является истинным периодом, во втором случае – нет .

Метод Касиски

Данный метод основан на том, что если гамма локально периодическая функция, то две одинаковые m-граммы открытого текста, отстоящие друг от друга на расстояние, кратное периоду гаммы, будут одинаково зашифрованы в некоторые одинаковые m-граммы, находящиеся на том же расстоянии друг от друга. Появление же одинаковых m-грамм в шифрованном тексте по другим причинам маловероятно. Следовательно, большинство расстояний между одинаковыми m-граммами делится на минимальный период. Поэтому на практике в качестве предполагаемого периода гаммы рассматривают наибольший общий делитель длин большинства расстояний между повторениями m-грамм. Эксперименты показали хорошую надежность этого метода, если в шифротексте имеются повторения триграмм и m-грамм при m, больше трех.

Библиотека БГУИР

Первый метод Фридмана

Первый метод Фридмана состоит в том, что для данного шифротекста B вычисляют индекс совпадения $IC(B)$ и сравнивают его с величинами

$$\frac{N-d}{d(N-1)}\sum_{i\in I}P_i^2+\frac{N(d-1)}{(N-1)d}\frac{1}{|I|}, d=1,2,3,...$$

При достаточной близости $IC(B)$ к одной из этих величин при некотором d предполагают, что период равен этому d.

С точки зрения статистической обоснованности, величину $IC(B)$ можно сравнивать с выражением

$$M_U(IC(B))=\frac{(l+1)lr+l(l-1)(d-r)}{N(N-1)}\sum_{i\in I}P_i^2+(1-\frac{(l+1)lr+l(l-1)(d-r)}{(N-1)N}\frac{1}{|I|},$$

$d=1,2,3,...,$

полученным для $N=ld+r$.

Метод характеризуется слабой эффективностью, что объясняется тем, что математические ожидания индекса совпадения шифрованного текста для класса U периодических гамм фиксированного периода d совпадают со значениями аналогичного параметра для целого ряда различных классов гамм.

Второй метод Фридмана

Этот метод основан на вычислении индекса совпадения, который состоит в опробовании возможных периодов следующей схемы. Для предполагаемого периода d выписывают d подпоследовательностей

$$\begin{array}{cccc} b_1 & b_{1+d} & b_{1+2d} & ... \\ b_2 & b_{2+d} & b_{2+2d} & ... \\ ... & ... & ... & ... \\ b_d & b_{d+d} & b_{d+2d} & ... \end{array}$$

Для каждой подпоследовательности подсчитывается ее индекс совпадения. Если все индексы совпадения в среднем близки к значению

$$\frac{1}{d}\sum_{i\in I}P_i^2,$$

т.е. к среднему значению индекса совпадения случайных шифротекстов, полученных с помощью гамм периода 1, то принимают величину d за истинный период, в противном случае опробывают другую величину периода.

Метод БШ

Предлагаемый метод определения периода гаммы по известному шифротексту B состоит в следующем.

Выписываются все пары номеров j, j', для которых $b_j=b_{j'}$. Пусть П(B) – множество таких пар. Очевидно, $|П(B)|=\sum_{i\in I}F_i(F_i-1)$, где F_i – частота встречаемости буквы i в шифротексте B.

Каждой паре (j, j') из П(B) ставится в соответствие расстояние $r(j,j')$, равное абсолютной величине разности между j и j'. Ищется максимальное по

Библиотека БГУИР

мощности подмножество П(B, d) пар в П(B), такое, что их расстояния $r(j, j')$ имеют некоторый общий наибольший делитель d, отличный от 1. Подсчитывается величина

$$ИБШ(B,d) = \frac{|П(B,d)|}{N(N-1)}$$

и сравнивается с величиной

$$M_U(ИБШ(B,d)) = \frac{(l-1)lr + l(l-1)(d-r)}{N(N-1)} \sum_i P_i^2 ,$$

где l, r определены равенством $N = l\,d+r$. Если эти величины близки, принимается гипотеза о том, что шифрование проводилось гаммой периода d.

3. КРАТКОЕ ОПИСАНИЕ ОСНОВНЫХ ОПЕРАТОРОВ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АЛГОРИТМОВ КРИПТОАНАЛИЗА

Программное обеспечение алгоритмов криптоанализа написано в среде MAPLE. Программа состоит из набора подпрограмм базовых алгоритмов криптопреобразований. Ниже приводится перечень основных подпрограмм.

Подпрограммы криптоанализа моноалфавитных шифров:

freqCounter – вычисляет частотную диаграмму заданного текста;

caesarrule – выполняет шифропреобразование последовательности по алгоритму Цезаря;

monoalph – выполняет преобразования букв текста в цифры;

encodemonoalph – шифрует открытый текст в соответствии с ключом и алгоритмом шифропреобразования;

caesarkeyfinder – определяет ключ шифра Цезаря по двум буквам соответственно упорядоченной частотной диаграммы шифротекста и вариационного ряда вероятностей открытого текста;

multrule – выполняет мультипликативное шифропреобразование;

maxcorr – выбирает предполагаемый ключ шифрования по максимальному значению корреляционной функции упорядоченной частотной диаграммы шифротекста и вариационного ряда вероятностей открытого текста;

multkeyfinder – определяет ключ key шифра, для этого представляет криптопреобразование как Y = key*X, после чего вычисляет ключ key = Y/X, при этом программа находит обратное число или множество предполагаемых обратных чисел по модулю N;

affinerule – выполняет аффинное шифропреобразование;

affinekeyfinder – вычисляет предполагаемый ключ аффинного шифропреобразования по двум буквам частотной диаграммы шифротекста и соответствующим им двум буквам вариационного ряда вероятностей открытого текста. Афинный шифр

Библиотека БГУИР

представляет собой комбинацию аддитивного и мультипликативного шифров. Ключ шифрования может быть определен, используя зависимость AX + B . Предположим, что имеется два символа шифротекста Y1 = A*X1 +B и Y2 = A*X2 + B. Можно вычислить ключи шифрования A и B, используя X1, X2, Y1,Y2 следующим образом. Вычислим (Y1 – Y2) = A*(X1 – X2), после чего определим A = (Y1 – Y2)/ (X1 – X2). При этом B = Y1 – A*X1. Заметим, что если (X1 – X2) делит 26, то имеется несколько решений;

invertaffinekey – вычисляет ключ расшифрования для аффинного шифра по предполагаемому ключу шифрования.

Подпрограммы криптоанализа многоалфавитных шифров:

vigen_enc – выполняет шифрование открытого текста по правилу Виженера;

vigen_dec – расшифровывает текст по правилу Виженера;

index_of_coincidence – вычисляет индекс совпадения;

freq, subfreq, sortfreq, topfreq – вычисляют упорядоченную частотную диаграмму заданного текста;

guesskey – проверяет гипотезу о длине ключа шифрования;

friedman, friedman2 – выполняют алгоритмы Фридмана по проверке гипотезы о длине ключа;

kasiski – выполняет алгоритм Касиски;

keylist – формирует множество возможных ключей;

keyletters – выбирает возможные ключевые знаки;

caesar –шифрует текст по правилу Цезаря.

4. СОДЕРЖАНИЕ И ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

4.1. Содержание работы

4.1.1. Изучить принцип построения методов криптографического анализа алгоритмов моно- и многоалфавитного шифрования.

4.1.2. Изучить основные операторы программ криптоанализа.

4.1.3. Составить программы алгоритмов криптоанализа моноалфавитных шифров и провести их моделирование с помощью программного пакета Maple по заданным открытым и шифрованным текстам.

4.1.4. Составить программы алгоритмов криптоанализа многоалфавитных шифров и провести их моделирование с помощью программного пакета Maple по заданным открытым и шифрованным текстам.

4.1.5. Провести анализ полученных результатов моделирования.

4.1.6. Выполнить расчетную часть лабораторной работы.

Библиотека БГУИР

4.2. Порядок выполнения работы

4.2.1. *Моделирование и исследование алгоритмов криптоанализа моноалфавитных шифров*

4.2.1.1. Составить и отладить программу моделирования алгоритмов криптоанализа аддитивного, мультипликативного и аффинного шифров с использованием программного пакета Maple.

4.2.1.2. Построить упорядоченный вариационный ряд открытого текста.

4.2.1.3. Получить от преподавателя исходные шифрованные тексты.

4.2.1.4. Вычислить упорядоченные частотные диаграммы зашифрованных текстов.

4.2.1.5. Оценить предполагаемые ключи шифрования и расшифрования исследуемых текстов.

4.2.1.6. Попытаться расшифровать исследуемые тексты, используя в качестве критерия осмысленность дешифрованного текста. В случае неудачи дешифрования внести коррективы в алгоритм криптоанализа и повторить попытку дешифрования.

4.2.1.7. Провести анализ полученных результатов и оценить эффективность исследуемых методов криптоанализа.

4.2.2. *Моделирование и исследование алгоритмов криптоанализа многоалфавитных шифров*

4.2.2.1. Составить и отладить программу моделирования алгоритмов криптоанализа шифра Виженера с использованием программного пакета Maple.

4.2.2.2. Получить от преподавателя исходные шифрованные тексты.

4.2.2.3. Провести анализ шифрованных текстов методами Фридмана, Касиски и БШ.

4.2.2.4. Вычислить предполагаемые ключи шифрования и расшифрования исследуемых текстов.

4.2.2.5. Используя результаты, полученные в п.3, попытаться расшифровать текст, используя в качестве критерия получение осмысленного текста. В случае неудачи дешифрования внести коррективы в алгоритм криптоанализа и повторить попытку дешифрования.

4.2.2.6. Провести анализ полученных результатов и оценить эффективность исследуемых методов криптоанализа.

Библиотека БГУИР

4.3. Расчетная часть

i. *Исходные данные.* Перехваченное сообщение имеет вид
“DXM SCE DCCUVGX ”.

Проверяется гипотеза, что сообщение составлено с помощью аффинных диграфов в 30-значном алфавите английского языка, в котором:

- буквы A–Z имеют численные эквиваленты 0–25;
- пробел=26,
- ?=27;
- !=28?,
- ‘ =29 (табл. 3).

Таблица 3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
R	S	T	U	V	W	X	Y	Z		?	!	“				
17	18	19	20	21	22	23	24	25	26	27	28	29				

Частотный анализ показал, что наиболее часто встречающимися диграфами в ранних шифротекстах были “M пробел”, “U пробел” “IH”. Предположим, что в английском тексте наиболее часто повторяющимися диграфами являются “E пробел”, “S пробел” и “пробел T”.

Задание:

а) найти ключи дешифрования и прочесть перехваченное сообщение;

б) найти ключи шифрования и зашифровать сообщение “YES I’M JOKING!”

4.3.2. *Исходные данные.* Перехвачено сообщение «ЦНТИ»

Проверяется гипотеза, что сообщение составлено с помощью аффинных диграфов в 33-значном алфавите русского языка (табл. 4).

Таблица 4

А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	

Частотный анализ ранних шифротекстов показал, что наиболее часто встречающимися диграфами были “ЦЯ”, “ЫТ”. Предполагается, что в русском языке наиболее часто повторяющимися диграфами являются “НО”, “ЕТ”.

Задание*:*

а) найти ключи дешифрования и прочесть перехваченное сообщение;

Библиотека БГУИР

б) найти ключи шифрования и зашифровать сообщение “Профиль защиты”.

5. СОДЕРЖАНИЕ ОТЧЕТА

5.1. Формулировка цели работы.
5.2. Выбор и обоснование выбранных методов криптоанализа.
5.3. Схемы алгоритмов криптоанализа.
5.4. Результаты криптоанализа и дешифрования.
5.5. Оценки эффективности исследуемых алгоритмов криптоанализа.
5.6. Расчетная часть.
5.7. Выводы.

6. КОНТРОЛЬНЫЕ ВОПРОСЫ

6.1. Пояснить сущность методов криптоанализа моноалфавитных шифров.
6.2. Пояснить сущность методов криптоанализа многоалфавитных шифров на примере шифра Виженера.
6.3. Почему алгоритм криптопреобразования Виженера можно рассматривать как гаммирующий шифр?
6.4. Как связана структура шифра Виженера и шифра Цезаря?
6.5. Почему применение диграфов повышает эффективность систем шифрования?
6.6. Какими характеристиками оценивается уровень защищенности шифров гаммирования?
6.7. Проведите сравнительный анализ методов Фридмана, Касиски и БШ.
6.8. Проведите сравнительный анализ уровня защищенности аддитивного, мультипликативного и аффинного шифров.
6.9. Как мощность алфавита влияет на эффективность криптоанализа аффинного шифра?

Библиотека БГУИР

ЛИТЕРАТУРА

1. Бабаш А.В., Шанкин Г.П. Криптография / Под ред. В.П. Шерстюка, Э.А. Применко. – М.: СОЛОН-3, 2002. – 512 с.

2. Иванов М.А. Криптографические методы защиты информации в компьютерных системах и сетях. – М.: КУДИЦ-ОБРАЗ, 2001. – 368 с.

3. Введение в криптографию / Под общ. ред. В.В. Ященко. 2-е изд., испр. – М.: МЦНМО «ЧеРо», 1999.

4. Menezes A.J., Van Oorschot P., Vanstone S. Handbook of Applied Cryptography. – N.Y.: CRC Press, 1996. – 780 p.

5. Бейкер А. Введение в теорию чисел / Пер с англ. Э.И. Ковалевской; Под ред. В.И. Берника. – Мн.: Выш. шк., 1995. – 127 с.

6. Говорухин В. И., Цибулин В. Г. Введение в Maple. Математический пакет для всех. – М.: Мир, 1997. – 208 с.

Библиотека БГУИР

Св. план 2003, поз. 59

Учебное издание

Саломатин Сергей Борисович

ЗАЩИТА ИНФОРМАЦИИ

Методическое пособие
к лабораторной работе
«Криптоанализ алгоритмов защиты информации»
для студентов специальностей
39 01 01 «Радиотехника», 39 01 02 «Радиоэлектронные системы»
дневной формы обучения

Редактор Н.А. Бебель
Корректор Е.Н. Батурчик

Подписано в печать 27.05.2003. Формат 60×84 1/16. Бумага офсетная.
Печать ризографическая. Гарнитура «Таймс». Усл. печ. л. 1,395.
Уч.-изд. л. 1,2. Тираж 150 экз. Заказ 42.

Издатель и полиграфическое исполнение:
Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Лицензия ЛП № 156 от 30.12.2002.
Лицензия ЛВ № 509 от 03.08.2001.
220013, Минск, П. Бровки, 6.

Библиотека БГУИР