Министерство образования и науки Российской Федерации
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Ю.А. КОТОВ

КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ

ШИФРЫ

Утверждено
Редакционно-издательским советом университета
в качестве учебного пособия

НОВОСИБИРСК
2016

УДК 004.056.55
К 736

Рецензенты:
д-р техн. наук, проф. *В.И. Гужов,*
д-р техн. наук, проф. *Е.В. Рабинович*

Работа подготовлена на кафедре защиты информации
для студентов III курса АВТФ дневной формы обучения

Котов Ю.А.
К 736 Криптографические методы защиты информации. Шифры: учебное пособие / Ю.А. Котов. – Новосибирск: Изд-во НГТУ, 2016. – 59 с.

ISBN 978-5-7782-2959-4

Представлены основные симметричные шифры в их взаимосвязи, теоретические и прикладные свойства данной группы криптографических методов защиты информации, а также способы применения этих методов.

Предназначено для студентов, обучающихся по направлению 10.03.01 «Информационная безопасность», специальности 10.05.03 «Информационная безопасность автоматизированных систем».

УДК 004.056.55

ISBN 978-5-7782-2959-4

ВВЕДЕНИЕ

Криптографические методы защиты информации – это формализованная группа методов защиты информации, имеющая собственную *количественную меру защищенности информации*. Изучением данных методов занимается наука *криптология*, подразделяющаяся на *криптографию и криптоанализ*. К криптографии относятся вопросы шифрования и расшифрования информации с использованием *ключей*, к криптоанализу – методы вскрытия шифров без знания ключа. Под *шифром* понимается совокупность обратимых преобразований открытых данных на множество защифрованных данных, задаваемых ключом и алгоритмом преобразования. Ключ – это конкретное *секретное* состояние параметров преобразования, обеспечивающее выбор только одного варианта из всех возможных для данного алгоритма.

Фундамент современной криптологии был заложен К. Шенноном в его работе середины прошлого века «Теория связи в секретных системах» [1].

Само же шифрование известно с незапамятных времен. В частности, обзор по истории криптографии в России можно найти в работе [16]. В более общей постановке криптографическая защита информации подразумевает решение следующих задач: разработка и анализ шифров, практическое применением шифров непосредственно для шифрования информации, а также применение криптографических методов в других областях зашиты информации, например, для идентификации/аутентификации данных и пользователей и т. д.

Для разработки и анализа шифров применяются самые разные математические методы из различных математических дисциплин: дискретной и вычислительной математики, теории множеств, теории чисел, теории алгоритмов, теории информации и кодирования, прикладной статистики и т. д. Ознакомиться с некоторыми из них применительно к криптографии и криптоанализу можно в работе [5], а также [11] и [4, 7, 8, 13–16].

В настоящее время существует большое количество стандартных и доступных алгоритмов шифрования [10, 12, 13, 16]. Применение шифров стало действительно массовым как в экономике, так и в личных целях. Сейчас основным вопросом при использовании криптографических методов защиты информации становится не разработка новых шифров и даже не криптоанализ – они все-таки остаются уделом узких специалистов, а массовое применение шифров широким кругом пользователей.

Массовый неподготовленный пользователь, применяющий стандартный шифр, обращается с ним как стандартный потребитель, и возможный провал такого шифрования несложно предугадать.

Специалистам известно, что результат шифрования всегда зависит от трех элементов: открытый текст, ключ, шифр, а значит, от того, как сбалансировал эти элементы пользователь шифра при его применении.

Именно знания и умения определять такой баланс в каждом конкретном практическом случае необходимы для того, чтобы сформировать из потребителя шифра профессионального пользователя. Для этого важно взаимосвязанное изложение вопросов теоретического и прикладного использования шифров такое, чтобы пользователь сам мог определить, когда и какой шифр или ключ является «хорошим» или «плохим», какое применение шифра является «правильным», а какое «неправильным», а также мог применять эти знания в конкретной практической обстановке, не прибегая к решению сложных математических задач и использованию объемной специализированной литературы. Таким вопросам в той или иной степени посвящены все работы [1–16], но можно дополнительно отметить [2–3, 6, 9, 12–13].

В настоящем учебном пособии мы будем последовательно формировать углубленное и в то же время простое и ясное представление о шифрах, шифровании и расшифровании, анализе и применении шифров, в первую очередь направленное на решение задачи профессионального использования шифров – как индивидуального, так и в рамках системы или организации.

Учебное пособие включает в себя пять разделов, каждый из которых содержит список контрольных вопросов, а первые четыре раздела – задания на лабораторный практикум. Задания могут выполняться с поддержкой распространенных программ: редакторов текстов и табличных процессоров типа MS Word и MS Excel, и/или простых программ, самостоятельно разработанных пользователем или полученных на кафедре защиты информации университета.

1. ШИФРЫ И ТЕКСТЫ

Шифром называется совокупность обратимых преобразований множества открытых данных на множество зашифрованных данных, задаваемое ключом и алгоритмом преобразования. Такие преобразования называются криптографическими.

Ключ – это конкретное секретное состояние некоторых параметров криптографического преобразования данных, обеспечивающее выбор только одного варианта из всех возможных для данного алгоритма.

Криптографические преобразования имеют две цели: обеспечить сокрытие *смысла* сообщения от лиц, не имеющих ключа; обеспечить обнаружение искажения исходной информации.

Мера качества криптографических преобразований называется *стойкостью шифра*. Она определяет возможность вскрытия шифра без знания ключа методами криптоанализа.

Шифры могут применяться для закрытия любого типа сообщений, передаваемых по каналам связи. Традиционным и значимым для всех сообщений объектом приложения шифров являются текстовые сообщения.

Основы современной теории шифров заложил К. Шеннон в работе [1].

1.1. Шифрование, кодирование и тайнопись

Установим различие между тремя способами преобразования информации: кодированием, шифрованием, тайнописью.

Очевидно, что защите подлежит информация, имеющая семантику (смысл) и ценность, выраженная на живом естественном языке. При всей кажущейся простоте этого утверждения из него следует важный практический вывод: не обязательно шифровать «все подряд». Более того, в отдельных случаях это совершенно недопустимо.

Например, есть таблица с результатами некоторого эксперимента (табл. 1). Набор цифр из таблицы: 601.1, …, 12.05, 150, …, 129, 31.4, …, 2,3 – без заголовков столбцов и порядка строк никакого смысла не имеет, как и шифрование этого набора. С другой стороны, если шифрованию подверглась общеизвестная и доступная информация, например стандартный заголовок документа или файла, то вместе с шифротекстом оказывается доступным и открытый текст, облегчающий раскрытие ключа шифрования.

Таблица 1

№ п/п	Скорость, км	Вес, кг	Разброс
1	601.1	150	31.4
…	…	…	…
112	12.05	129	2,3

Кодирование устанавливает взаимооднозначное соответствие между сигналом, несущим информацию о реальном физическом объекте, и его кодом. Следовательно, каждый раз, когда повторяется исходный объект, повторяется и его код. Закодированное сообщение полностью сохраняет семантику и свойства исходного сообщения.

Цель кодирования – сохранить максимально полно и точно информацию об объектах кодирования в процессах ее передачи, хранения и воспроизведения.

Рассмотрим сообщение на естественном языке (ЕЯ). При кодировании сообщения на естественном языке в качестве элементарных исходных объектов для письменных сообщений выступают символы алфавита. Для устных сообщений – фонетические единицы (фонемы) языка (рис. 1).

Фонемы устного языка при переводе их в письменную форму сначала кодируются знаками, т. е. *знаковым* или *символьным* кодом. Символы языка, как знаки фонем, в свою очередь, могут быть закодированы другим знаковым либо *числовым* кодом и т. д. (рис. 1).

Шифрование устанавливает *многозначное соответствие* между объектом и его (шифро-) кодами, которое становится однозначным только при наличии ключа. Одному и тому же исходному объекту при шифровании соответствуют разные (шифро-) коды.

Тайнопись – это невоспроизводимость части или всей информации при использовании стандартного для нее способа отображения. Например, если предполагается, что текст воспроизводится в лучах

отраженного белого цвета, то не будут воспроизводиться сообщения, нанесенные термочернилами, краска, проявляющаяся в особых частях спектра, и т. д.

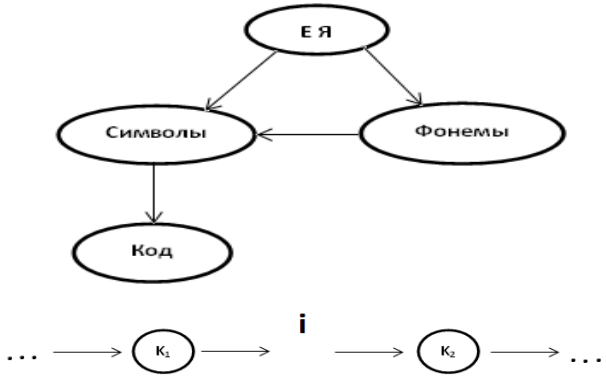

Рис. 1

Если стандартный способ чтения текста – последовательный просмотр текста слева направо и сверху вниз (или другой), то, например, анаграмма – это тайнопись относительно этого способа; другой пример – использование «решеток» для фиксированного текста, «25 кадр» – для кино- и видеофильмов. В современных компьютерных системах тайнопись часто использует запись информации в аудио-, видеофайлах и других форматах, не требующих битовой точности представления информации.

Примем следующие соглашения по использованию терминов.

1. Недопустимо использовать термин «кодирование» вместо термина «шифрование», только – «шифровать/расшифровать».

2. Использование термина «код», «шифрокод» применительно к отдельному символу оправданно. Применительно ко всему тексту следует использовать термин «шифротекст».

3. Для тайнописи не следует использовать термин «закрытие информации», так как на самом деле речь идет о сокрытии информации.

4. Будем говорить, что шифрование *вырождается* в кодирование, если между кодами открытого текста и шифротекста существует взаимооднозначное соответствие.

5. Будем говорить, что кодирование *вырождается* в копирование, если кодируемый объект и его код совпадают.

1.2. Частотные характеристики текстов

Хотя шифрование может применяться и к устным, и к письменным сообщениям (текстам), будем в качестве объекта применения шифров рассматривать только последние. Текст на естественном языке состоит из знаков, которые разделяются на знаки алфавита языка сообщения (называемые обычно буквами) и вспомогательные – пробел, цифры, знаки пунктуации и т. д.

Термин *знак*, используемый и самостоятельно, является основой и других терминов – *буква, символ, код*, понимаемых в разных дисциплинах по-разному. Так, в лингвистике термин *символ* имеет иное значение, чем термин *буква*, так как считается, что первый передает некоторую информацию («смысл»), а второй никакой смысловой информацией не обладает. В программировании одиночный *знак*, в отличие от чисел, обычно называют *символом*. Информатика оперирует термином *код*, в первую очередь как *код* (обычно числовой) *знака*, но может понимать под ним и сам *знак* в случаях использования не числовых, но *знаковых кодов*. В криптографии используют смешанную терминологию в зависимости от однозначности понимания термина в каждом конкретном случае.

Алфавит **А** *любого языка представляет собой множество упорядоченных кодов символов, обозначающих буквы этого языка*. Буквы языка однозначно связаны с их порядковым номером в исходном алфавите, но могут быть представлены разными, в том числе неизвестными, знаками, и в то же время одинаковые знаки в разных текстах могут обозначать одну и ту же, но возможно неизвестную букву. Будем далее под *символом* понимать букву языка, код которой (значение) в исходном алфавите нам может быть заранее неизвестен.

Отсюда следует, что буква языка на самом деле имеет значение, и это значение – порядковый номер (код) буквы в исходном алфавите. В текстах эта буква может быть представлена не одним, а разными знаками: заглавные и строчные буквы, прописные и печатные и т. д. Возможно даже, что буква имеет «размер», и представлена не одним, а некоторой последовательностью знаков (в лингвистике известны понятия *диграмма, дифраф* для таких случаев).

Будем для упрощения считать, что все буквы алфавита представлены одним одиночным знаком каждая. Тогда мощность алфавита языка N_A соответствует количеству букв алфавита и представляет начальную числовую характеристику для текстов на данном языке.

Будем также считать, что в текстах используется только один вспомогательный символ – пробел.

Математически это означает, что все слова текста представляют собой разделенные пробелом числа в N_A-ричной системе счисления, определяемой алфавитом **A**.

Первое, что можно определить по тексту **T** – это множество используемых в нем знаков **AT** и мощность этого множества N_{AT}, объем текста N_T как общее количество используемых в нем символов. Полагая, что на множестве **AT** существует некоторое упорядочение, мы можем говорить об алфавите текста **AT.** Сопоставляя значение N_{AT} с известными значениями мощности алфавитов различных языков, при определенных условиях можно сделать предположение о принадлежности **T** к конкретному языку. Для более точного определения языка текста можно использовать так называемый *индекс совпадения* (У. Фридман, 1920 г.) [6]. Он показывает вероятность того, что две случайно выбранные из текста буквы совпадают. Значения индекса для европейских языков приведены в табл. 2.

Таблица 2

Язык	Русский	Англ.	Франц.	Нем.	Итал.	Испан.
$I(x) \approx c$	0,0529	0,0662	0,0778	0,0762	0,0738	0,0775

Другие характеристики, которые можно получить из текста, – это число повторений сочетаний знаков в тексте – так называемых (k)-грамм. В общем случае это сочетание (k) элементов: знаков, слов, предложений и т. д. Но мы будем использовать термин (k)-граммы только для знаков и символов.

Тогда (1)-граммы (*униграммы*) – это число повторений одиночных символов текста, (2)-граммы (называемые *биграммами*) – число повторений сочетаний пар символов, (3)-граммы (*триграммы*) – число повторений трехсимвольных цепочек и т. д. Обычно их называют частотными характеристиками, так как их легко привести к частотной форме, нормировав к общему количеству символов текста или количеству определенных символов, но, как правило, чаще используют как число повторов.

Число повторений сочетаний знаков используется в различных методах частотного анализа текстов. Так, простое сопоставление числа повторений одиночных символов текста с известными эталонными значениями вероятности появления символов в текстах на некотором

языке, которое легко осуществить методом упорядочения, может дать начальное приближение при решении задачи восстановления простой замены букв.

Однако погрешность такого приближения достигает 0,7 по символам даже при объемах текста свыше 350 000 символов, т. е. можно считать, что мы никогда не получим более 30 % правильных букв при таком приближении. Более точную информацию о тексте можно получить, рассмотрев частотные характеристики высокого порядка, например, символьные биграммы.

Пусть, например, для русскоязычного текста, представленного в простой замене символов, у нас есть таблица биграмм, симметрично упорядоченная по числу появления символов в тексте. На основании этой таблицы при достаточном объеме текста можно идентифицировать следующие символы текста [8] (рис. 2 и 3).

1. Если разность сумм по строке первых семи значений и остальных меньше нуля, то данный символ, скорее всего, гласный.

2. Первые четыре символа таблицы, для которых суммарное количество ненулевых биграмм по строке и столбцу больше, чем для других символов, представляют символы (О, Е, А, И).

3. Результат п. 2 сопоставляется с п. 1 для проверки, и символы, выделенные по критерию п. 2, переносятся (в случае необходимости) в начало таблицы с сохранением исходного взаимного упорядочения.

4. По столбцам таблицы выделяются первые два символа, для которых сумма первых четырех значений по столбцу равна нулю – это символы (Ь, Ы), или наоборот.

5. По строкам таблицы выделяется первый символ, для которого сумма первых четырех значений по строке равна нулю – это символ (Э).

6. Начиная с 5-го символа, в таблице выделяется подматрица размером 10×10. Максимальное значение в этой подматрице определяет биграмму (и соответственно символы) (СТ).

7. Для выделенных ранее символов (О, Е, А, И) максимальное значение (значения) в строке будет появляться, скорее всего, для символа (О). Возможно, это значение будет идентифицировать биграмму (ОВ), возможно (ОР).

8. Максимальное значение в столбце ранее идентифицированного символа (Ь) (без учета ранее идентифицированного символа «Т»), скорее всего, указывает биграмму (ЛЬ), т. е. символ (Л).

9. Среди всех диагональных значений наибольшее, скорее всего, указывает на биграмму (НН) – символ (Н).

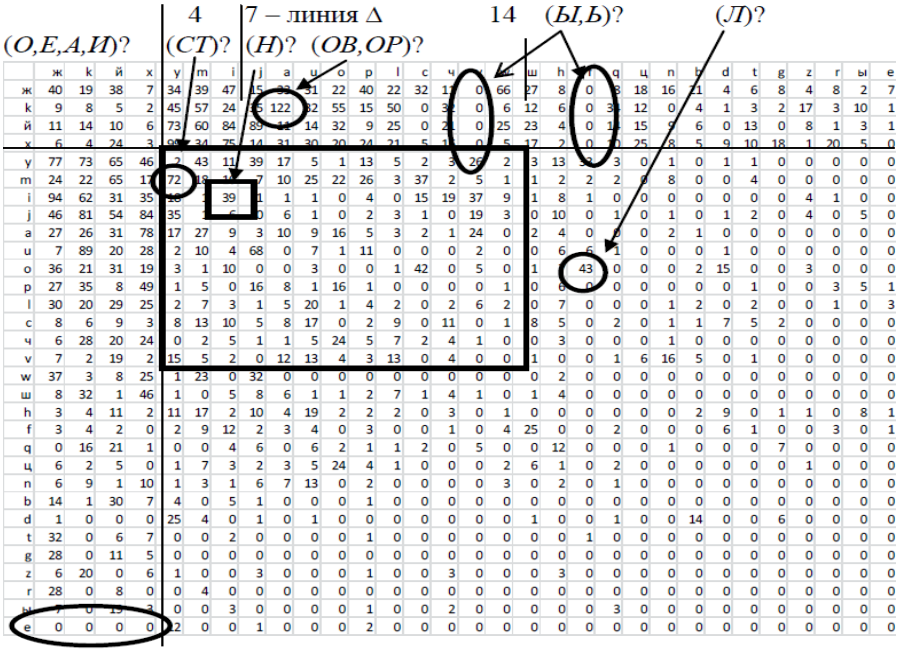

Рис. 2

Таблица биграмм

D
d_1
…
d_{31}

	$\boldsymbol{s_1}$		$\boldsymbol{s_{31}}$
$\boldsymbol{s_1}$	$b_{1,1}$		$b_{1,31}$
…	…		…
$\boldsymbol{s_{31}}$	$b_{31,1}$		$b_{31,31}$

$\sum^1$
$m_1^{\ 1}$
…
$m_{31}^{\ 1}$

$\mathbf{K_1}$	$\mathbf{K_2}^{\mathbf{T}}$	$\sum^3$
$k_1^{\ 1}$	$k_1^{\ 2}$	$r_1 = k_1^{\ 1} + k_1^{\ 2}$
…	…	…
$k_{31}^{\ 1}$	$k_{31}^{\ 2}$	$r_{31} = k_{31}^{\ 1} + k_{31}^{\ 2}$

Δ

$\sum^2$	$m_1^{\ 2}$	…	$m_{31}^{\ 2}$

$\mathbf{K_2}$	$k_1^{\ 2}$	…	$k_{31}^{\ 2}$

Специальная вставка:

Значения + транспонировать

Рис. 3

Оценки пп. 1–8 следует получать, предварительно обнулив значения диагональных биграмм в таблице и выделив их в отдельный вектор **D** (рис. 3). Все расчеты несложно проделать с помощью любого табличного процессора, в частности MS Excel, рекомендуемая начальная форма таблиц в котором предложена на рис. 3.

Использование частотных характеристик биграмм дает более точную, чем одиночные символы, но все же приближенную информацию о тексте. Для дальнейшего повышения точности анализа текста, особенно при уменьшении его объема (что характерно для криптографии), необходимо использовать частотные характеристики все более высоких порядков, в том числе частотные словари языков.

Пояснения к обозначениям, принятым на рис. 3:

$$d_i = b_{i,i},\ \ b_{i,i} = 0;$$

$$m_i^1 = \sum_{j=1}^{31} b_{i,j},\ \ m_i^2 = \sum_{j=1}^{31} b_{i,j};$$

$$k_i^1 = \text{счет, если } (b_{i,j} > 0,\ j = \overline{1,31}),\ k_i^2 = \text{счет, если } (b_{j,i} > 0,\ j = \overline{1,31});$$

$$\Delta_i = \sum_{j=1}^{7} b_{i,j} - \sum_{j=8}^{31} b_{i,j}.$$

Контрольные вопросы

1. Дайте определение шифра, ключа.
2. Чем шифрование отличается от кодирования?
3. Что такое тайнопись?
4. Для чего применяются шифры?
5. Что такое алфавит?
6. Какое значение имеет буква?
7. Что такое (*k*)-граммы?
8. Что такое индекс совпадения и как он вычисляется?
9. Как в закодированном тексте выделить с помощью биграмм:
   а) символы (О, Е, А, И);
   б) символы (Ь, Ы)
   в) символ (Э);

г) символы (С, Т);
д) символ (Л);
е) символ (Н);
ж) символы (О, В)?

Задание на лабораторный практикум

«*Частотные характеристики открытого текста*»

Задан один последовательный текст на одном естественном языке. Текст разбит на следующие друг за другом и неперемешанные фрагменты (варианты). Произведение не художественное. Текст содержит только символы (буквы) одного алфавита и символ «пробел» (цифры, символы пунктуации и специальные символы исключены). Для кодирования текста использовались четыре различных знаковых кода, различающихся множествами используемых в них символов (знаков). Каждый фрагмент (вариант) текста закодирован только одним из кодов.

1. Выбрать один фрагмент исходных данных.
2. Определить язык исходного сообщения – русский или английский.
3. Восстановить исходный текст на языке сообщения, пользуясь материалами раздела 1.2 и работами [6, 8].

2. ШИФРЫ ПЕРЕСТАНОВКИ ВЗАИМОСВЯЗЬ ПЕРЕСТАНОВОК

Шифрами с *симметричным ключом* называются шифры, использующие для шифрования и расшифрования один (и тот же) ключ. Для реализации секретной связи с симметричным ключом среди абонентов системы до начала обмена зашифрованными сообщениями должен быть распространен один и тот же ключ. Симметричные шифры представлены двумя классами шифров: шифрами *перестановки* и шифрами *замены* (называемые иногда *подстановкой*).

Все шифры перестановки переставляют символы шифруемого текста в пределах блока текста фиксированного размера m по одному ключу (рис. 4).

T	=>	$\mathbf{T_1}$	$\mathbf{T_2}$	$\mathbf{T_3}$	...	$\mathbf{T_n}$
$n \times m$		m	m	m		m

Рис. 4

Известно большое количество различных шифров перестановки [12,13], которые имеют в настоящее время в основном историческое и методическое значение. Все «классические» шифры были созданы в докомпьютерную эпоху, когда необходимо было повысить эффективность трудоемкого процесса ручного шифрования наряду со стойкостью шифра. В современных условиях прикладное значение данные шифры имеют только в тех случаях, когда необходимо провести шифрование без помощи компьютера, тем более что все «классические» перестановки могут быть сведены к одному шифру перестановки.

Определим шифр перестановки линейного блока, или просто *линейную перестановку* (ЛП), по заданному ключу. Пусть задан линейный блок текста длиной в m символов, и ключ размером m, показывающий, на какое место в блоке шифротекста размером m помещается текущий символ открытого текста из очередных m символов (рис. 5).

Открытый **T**	1	2	3	...	m
Ключ	**1**	**12**	**8**	**...**	***k***
Шифротекст **T**	1	12	8	...	k

Рис. 5

Здесь символы исходного текста представлены их порядковым номером в линейной последовательности блока исходного текста, $k<m$. Шифротекст формируется последовательным считыванием символов блока шифротекста размером m.

Введем понятие *функциональной эквивалентности* шифров. Пусть $\mathbf{E}_i$ – шифр i, $\mathbf{K}_i$ – множество всех его ключей, **T** – множество открытых данных.

Будем называть шифр $\mathbf{E_1(K_1,T)}$ *функционально-эквивалентным* шифру $\mathbf{E_2(K_2,T)}$, если для любого k_i^1 из $\mathbf{K_1}$ существует k_i^2 из $\mathbf{K_2}$ такой, что результаты шифрования любого открытого текста t_j из **T** шифрами $\mathbf{E_1}\left(k_i^1, t_j\right)$ и $\mathbf{E_1}\left(k_i^2, t_j\right)$ совпадают.

Утверждение 1. Все шифры перестановки функционально эквивалентны шифру перестановки линейного блока по ключу.

Покажем справедливость данного утверждения для основных способов перестановки, реализованных в «классических» шифрах. Дополнительно получим возможность непосредственно изучить и сами исходные шифры.

2.1. Простая перестановка

Определяется двумерный блок (матрица, таблица) размерностью $n \times m$. Блоки исходного текста размером $n \times m$ записываются в данную таблицу, например по строкам (можно и по столбцам), а затем считываются по столбцам (во втором случае – по строкам), формируя блок шифротекста размерностью $n \times m$.

Обозначим каждый символ текста его порядковым номером в тексте. Схема алгоритма выглядит так, как показано на рис. 6.

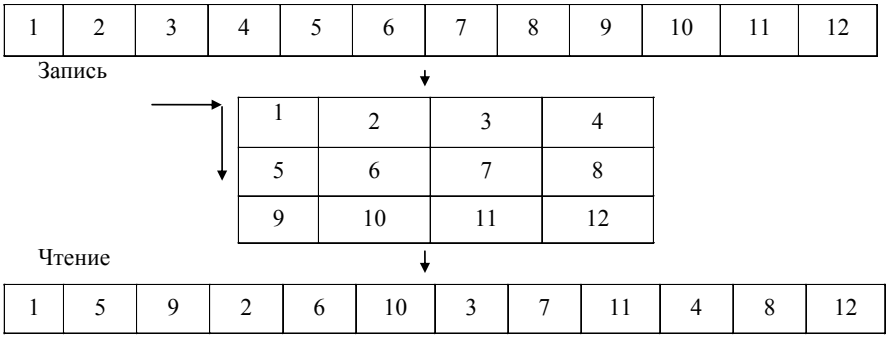

Рис. 6

На рис. 6 верхняя строка представляет собой линейную последовательность номеров символов блока исходного текста длиной $n \times m$, промежуточная таблица – заданный ключ размерностью $n \times m$ и его применение к исходному тексту, заключительная строка – ключ ЛП размерностью $n \times m$, необходимый для функционально-эквивалентной замены исходной перестановки. Аналогично интерпретируются и рис. 7–21.

Полезное замечание: при шифровании перестановкой для каждого блока исходного текста можно выделить свой алфавит.

Пример: НОВО СИБИ РСКИ Й
(В,О,Н) (Б,И,С) (И,К,С,Р) (Й).

Замена исходного шифра шифром ЛП позволяет наглядно представить наличие или отсутствие функциональной зависимости элементов ключа исходного шифра. Для этого определим *характеристическую функцию ключа* как интерполирующую его значения кусочно-линейую функцию.

График характеристической функции ключа ЛП раздела 2.1 (рис. 7) показывает сильную функциональную зависимость элементов ключа простой перестановки: четыре раза используется практически один и тот же ключ: 0; +4; +4; со сдвигом начального значения. Зависимость такова, что ключ простой перестановки может быть описан несложной рекуррентной формулой.

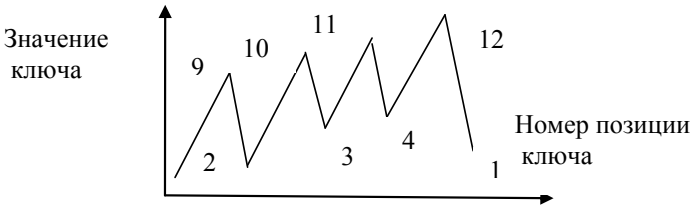

Рис. 7

Пример показывает фундаментальное значение для криптографии вопроса о роли ключа и его «длине» – из примера видно, что ключом является даже не номер позиции символа в тексте, а нечто другое. Что именно, еще предстоит разобраться.

2.2. Одиночная перестановка по ключу

Шифр простой перестановки раздела 2.1 дополняется ключом длиной n, определяющим порядок следования столбцов (строк) основного ключа $n \times m$, например, ключом (2, 1, 4, 3) для рассмотренной на рис. 6 простой перестановки (рис. 8).

График характеристической функции ключа рассмотренной одиночной перестановки по ключу (рис. 9) показывает, что функциональная зависимость элементов ключа хотя и ослаблена по сравнению с разделом 2.1, но незначительно.

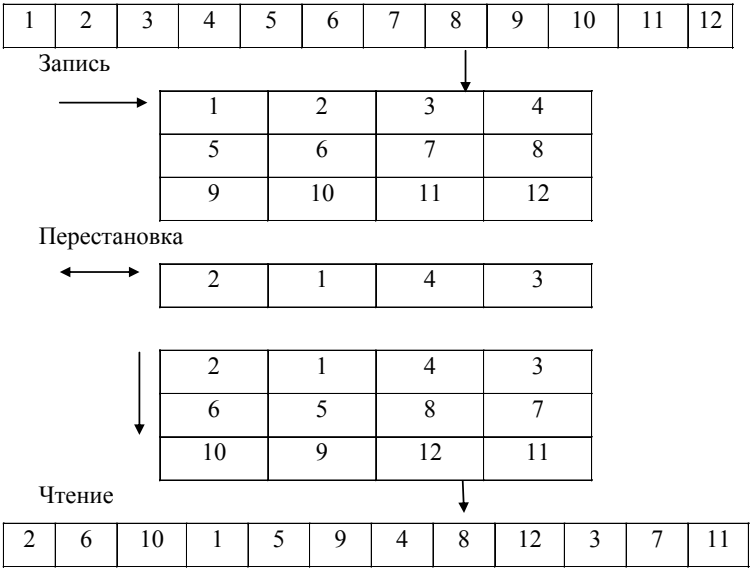

Рис. 8

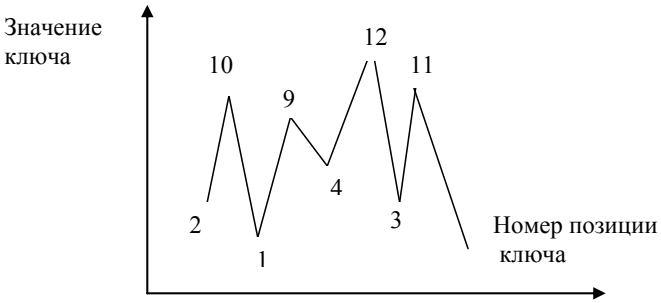

Рис. 9

2.3. Двойная перестановка по ключу

Шифр одиночной перестановки по ключу раздела 2.2 дополняется еще одним ключом размером m, определяющим порядок следования строк (столбцов) основного ключа $n \times m$ (рис. 10).

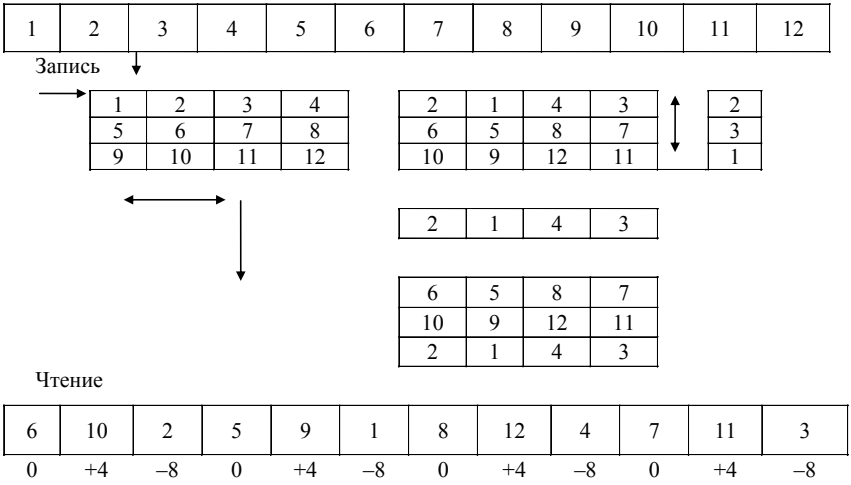

Рис. 10

Нижняя строка на рис. 10 показывает, что функциональная зависимость элементов ключа в значительной степени сохраняется.

2.4. Перестановка с запретом записи

Дополнительный ключ указывает ячейки основного ключа, запись в которые не производится. Например, запрет записи в ячейку (2,2) для шифра простой перестановки раздела 2.1 показан на рис. 11.

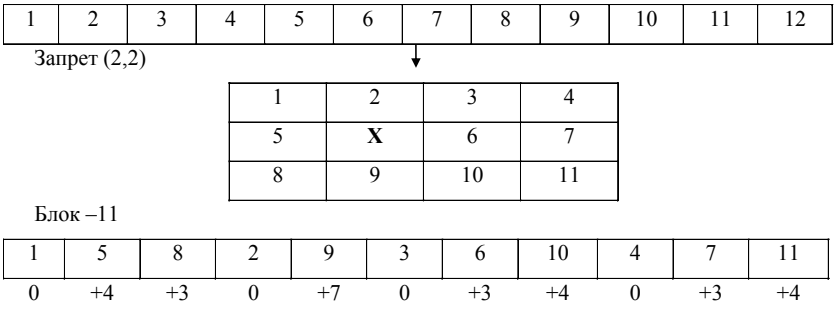

Рис. 11

Для этой перестановки характерны следующие изменения. Уменьшается размер блока – ($n \times m - p$), где p – количество запрещенных ячеек. Блок теперь не обязательно должен иметь матричную форму. Изменяется внутренний ключ: появляются разные периоды (длины) внутренних ключей, но для отдельных групп элементов ключа все еще заметна повторяемость.

2.5. Перестановка с использованием «магических» квадратов

«Магический» квадрат – это квадратная матрица $m \times m$ последовательных натуральных чисел от 1 до $m \times m$ такая, что сумма чисел по каждому столбцу, строке и диагонали одинакова. Перестановка с использованием магических квадратов – это простая перестановка с таблицей ключа $m \times m$ и вписанным в нее магическим квадратом. Символ отрытого текста с порядковым номером i помещается в клетку таблицы, содержащую число i. Далее – по стандартному алгоритму (рис. 12).

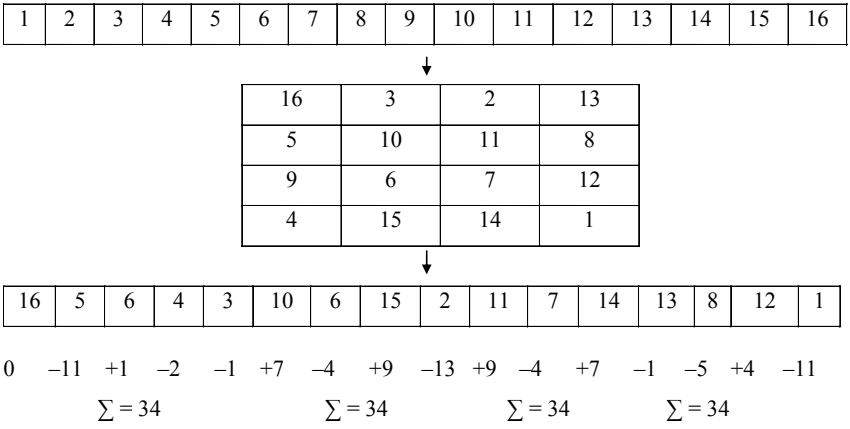

Рис. 12

Функциональная зависимость между элементами ключа рассматриваемой перестановки настолько ослабла, что практически исчезла. Если вместо магических квадратов применить произвольные матрицы $n \times m$ натуральных чисел от 1 до $n \times m$, то данная перестановка практически вплотную приближается к шифру ЛП.

2.6. Перестановка «по маршрутам»

Еще один шаг в направлении шифра ЛП. Можно встретить перестановку по маршрутам, проходящим по вершинам некоторого гиперкуба (гиперкуб размерности N имеет 2^N вершин), например, маршрутам обхода вершин без повторов. Представим структуру трехмерного гиперкуба (рис. 13). Тогда маршрут 1 представлен на рис. 14, маршрут 2 – на рис. 15.

5			6
	1	2	
	3	4	
7			8

Рис. 13

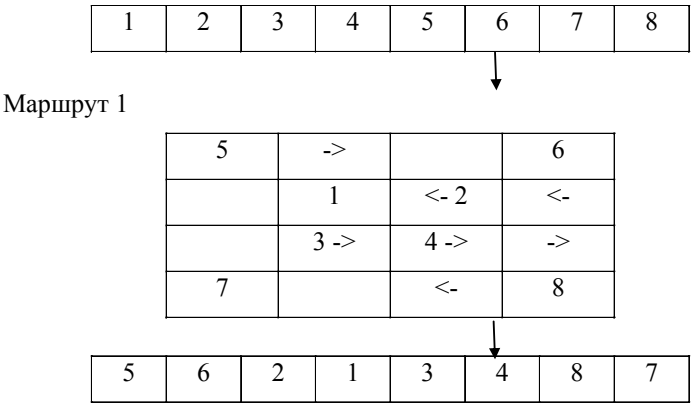

Рис. 14

На рис. 14 обход вершин осуществляется в порядке: 5, 6, 2, 1, 3, 4, 8, 7. Эта последовательность определяет ключ ЛП и порядок заполнения блока шифротекста символами открытого текста.

На рис. 15 обход вершин осуществляется в порядке: 5, 1, 3, 4, 2, 6, 8, 7, что определяет ключ ЛП и порядок заполнения блока шифротекста.

Как мы видим из рис.14 и 15, «маршрут» обхода вершин непосредственно является ключом линейной перестановки, единственным огра-

ничением для которого остается только размер блока, определяемый количеством вершин гиперкуба.

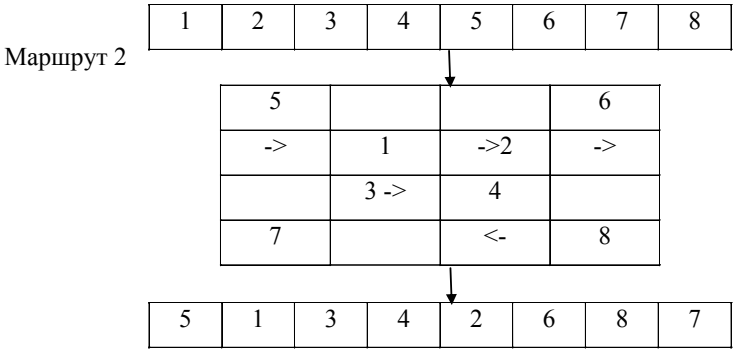

Рис. 15

2.7. Перестановка, использующая разные размеры блоков

Возьмем в качестве исходного шифр простой перестановки, но последовательно использующий два блока разных размеров: $n \times m$ и $k \times p$. Общая длина блока – $(n \times m + k \times p)$, рис. 16.

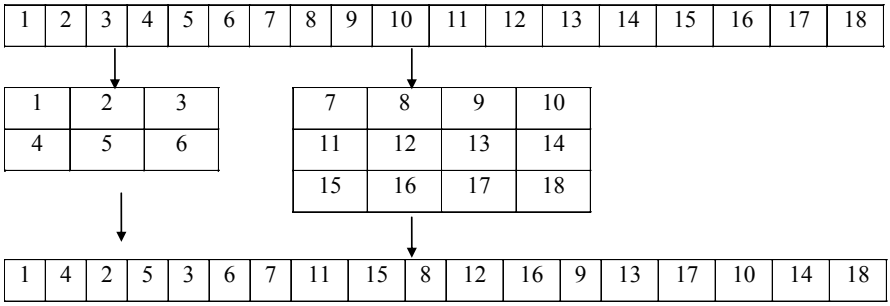

Рис. 16

2.8. Перестановка, использующая смену направления записи/чтения

Дополнительным ключом для простой перестановки может являться направление записи/чтения для каждого последующего блока, например: З, Ч, З, З, …. (рис. 17).

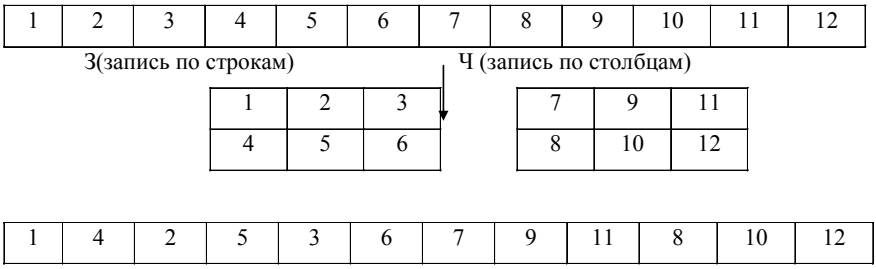

Рис. 17

Обратим внимание, что для квадратных матриц-ключей такая перестановка эквивалентна простой перестановке с одним ключом (рис. 18).

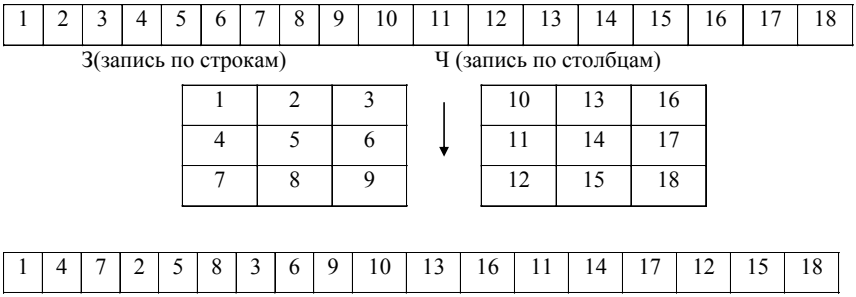

Рис. 18

2.9. Повторная перестановка и композиция перестановок

Это не самостоятельный алгоритм, а применение общего рекурсивного принципа: к зашифрованному тексту вновь применяется алгоритм шифрования (в данном случае – перестановка), но уже с другим клю-

чом, и т. д. Один шаг многократного циклического шифрования в криптографии обычно называется *циклом*, или *раундом*.

Покажем, что циклические перестановки также функционально эквивалентны шифру ЛП.

Пусть первая простая перестановка использует ключ 2×2 (рис. 19), вторая – 3×3 (вариант 1) или 3×2 (вариант 2) (рис. 20).

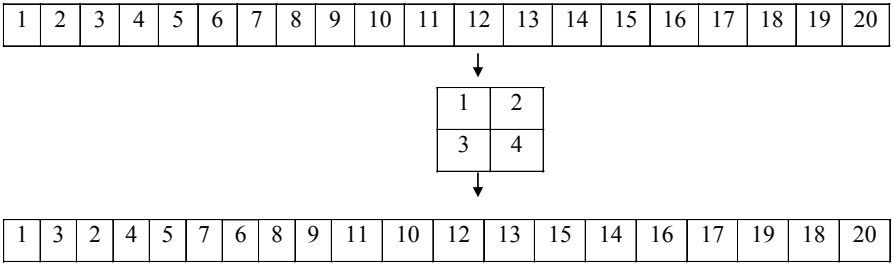

Рис. 19

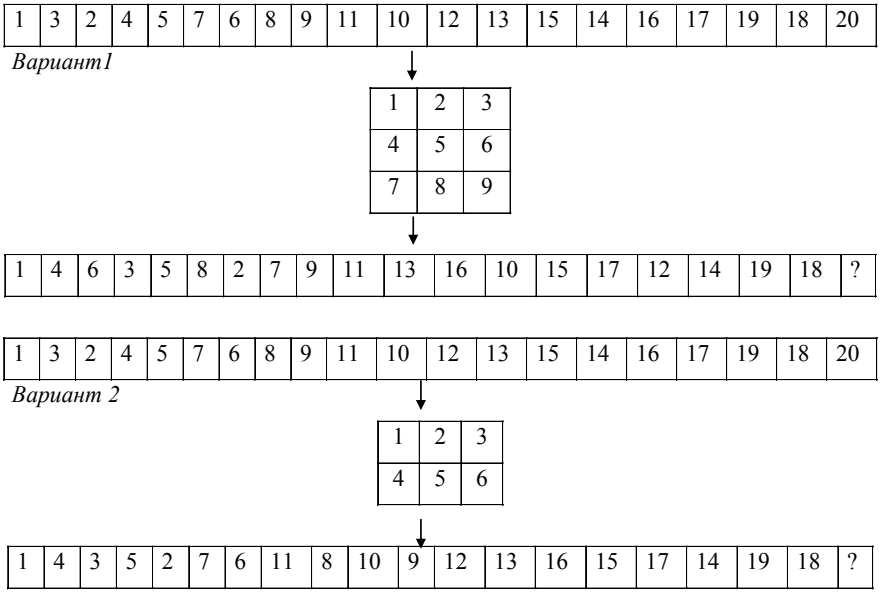

Рис. 20

Как видно из рис. 20, для функционально эквивалентной замены циклических перестановок одним шифром ЛП необходимо решить проблему выравнивания длин всех используемых в повторных перестановках блоков.

Общее решение этой задачи дает размер блока ЛП, равный наименьшему общему кратному (НОК) для всех размеров блоков, использовавшихся при циклическом шифровании, НОК: (2×2) × (3×3) = = 36 для варианта 1 и НОК: 4×6 = 24 для варианта 2.

Обратим внимание на следующую особенность повторных перестановок: если длины используемых блоков – это простые числа, то размер общего блока ЛП очень быстро возрастает: 2×5 = 10×7 = = 70×11 = 770×13 = 10 010 (всего за пять перестановок) и т. д.

Ключ для ЛП, заменяющий циклические перестановки, формируется представленным на рис. 21 способом.

Ключ 1-й перестановки (18)

18	3	6	…	4	1	…	17	5	…	10
1	2	3	…	6	7	…	12	13		18

Ключ 2-й перестановки (6)

6	3	2	…	1	6+6×1	…	1+6×1	6+6×2	…	1+6×2
1	2	3	…	6	7		12	13	…	18

1-й промежуточный ключ 1 (НОК: 18)

4	6	3	…	18	17	…	1	10	…	5
1	2	3	…	6	7	…	12	13	…	18

. . . . . .

Ключ *n-й* перестановки (18)

A_{k1}	A_{i1}	A_{j1}	...	A_{m1}	A_{n1}	...	A_{p1}	A_{x1}	...	A_{z1}
1	2	3	…	6	7	…	12	13	…	18

Окончательный ключ ЛП (НОК всех предыдущих ключей 18)

A_k	A_i	A_j	...	A_m	A_n	...	A_p	A_x	...	A_z
1	2	3	…	6	7	…	12	13	…	18

Рис. 21

Отметим также такую неприятную особенность перестановки, как возможность раскрытия текста при повторной перестановке.

На примерах рассмотренных шифров нетрудно увидеть, что «классические» шифры перестановок в большей степени ориентированы на подготовку ключа, чем на саму перестановку. Они направлены на то, чтобы сделать ключ как можно более трудно раскрываемым или случайным.

Контрольные вопросы

1. Какие шифры называются симметричными?
2. Какова особенность секретной связи при использовании симметричных шифров?
3. Какие шифры называются перестановкой?
4. Как определяется функциональная эквивалентность шифров?
5. В чем заключается единственность шифра перестановки?
6. Приведите схему перестановки:
   а) простой;
   б) одиночной по ключу;
   в) двойной по ключу;
   г) с запретом записи;
   д) с использованием магических квадратов;
   е) по «маршрутам»;
   ж) со сменой направления записи/чтения;
   з) использующей разные размеры блоков;
   и) повторной перестановку.
7. Как определить размер блока линейной перестановки для повторных перестановок?
8. В чем заключаются фундаментальные особенности шифров перестановки?
9. Что называется циклом или раундом при шифровании?
10. Что такое характеристическая функция ключа и для чего она может быть использована?
11. Какому шифру перестановки функционально эквивалентны остальные шифры перестановки?

Задание на лабораторный практикум

«*Шифр перестановки*»

Смоделировать шифр линейной перестановки для одного из указанных шифров.

1. Простая перестановка, блоки: 5×7, 4×9, 6×5.

2. Одиночная перестановка по ключу, блоки: 4×9, 5×7, 8×4. Дополнительный ключ формировать по своему имени.

3. Двойная перестановка по ключу, блоки: 6×5, 7×5, 10×3. Дополнительные ключи формировать по своим имени и отчеству.

4. Перестановка с запретом записи, блоки: 7×7, 6×7. Дополнительный ключ формировать по наиболее часто встречающейся букве своих фамилии, имени и отчества.

5. Перестановка с использованием «магического» квадрата, блоки: 6×6, 5×7, 5×5. Дополнительный ключ формировать на основе своих фамилии, имени и отчества.

6. Перестановка «по маршрутам», блок 32. Дополнительный ключ формировать на основе своих фамилии, имени и отчества.

7. Перестановка, использующая различные размеры блоков: (3×7, 4×3), (3×6, 5×4), (4×7, 4×5).

8. Перестановка, использующая смену направления записи/чтения, блоки: 3×4, 2×9, 5×3;

1-й блок – запись по строкам, 2-й блок – запись по столбцам, 3-й блок – запись по столбцам.

Провести повторную перестановку зашифрованного текста с помощью той же перестановки, но использующей другой ключ.

Смоделировать повторную перестановку с помощью шифра ЛП и сравнить результаты повторного и прямого шифрования, расшифровав с помощью полученного ключа ЛП повторно зашифрованный текст.

3. ДЛИНА И РАЗМЕР КЛЮЧА ШИФРА

При рассмотрении перестановок было показано, что метрика ключа связана не только с его размером. В криптографии длину ключа связывают не столько с его размером, сколько с понятием стойкости шифра.

Стойкостью шифра называется его *расстояние единственности* [1], которое определяет минимальный объем зашифрованного данным

шифром текста, при уменьшении которого никакими методами анализа невозможно однозначно восстановить исходный текст [1].

Можно получить лишь так называемые *правильные* варианты открытого текста, количество которых и определяет неоднозначность решения.

Следовательно, стойкость шифра – это допустимый объем информации, зашифрованный при использовании *одного* ключа. Измеряется стойкость шифра в тех же единицах, что и исходный текст, т. е. в битах.

Меру стойкости шифра как «расстояния единственности» ввел К. Шеннон [1]. Он рассмотрел на основе своей информационной меры – *энтропии* – взаимосвязь неопределенности открытого текста, ключа и шифротекста. Шеннон доказал, что для каждого языка сообщения, шифра и множества его ключей существует такое «расстояние единственности», измеряемое в единицах объема исходного текста, при превышении которого статистическими методами анализа можно вскрыть исходный текст, а при уменьшении – нельзя. При этом размер ключа (или его *период* – как период повторяемости) может быть и меньше длины открытого текста, измеряемого в тех же, что и ключ, единицах измерения.

Кроме того, необходимо также, чтобы ключ никак себя не проявлял в зашифрованном тексте, несмотря на все допустимые изменения в открытом тексте. Следовательно, длина ключа в контексте определения стойкости шифра не совпадает с его размером.

Но практичного способа определения длины ключа в терминах стойкости шифра для произвольного шифра не существует, хотя этому и была посвящена большая часть работы [1]. В результате были получены только оценки такой длины для некоторых шифров.

На практике чаще применяют *комбинаторную* (или *переборную*) меру стойкости шифра, получаемую как число всех возможных ключей данного размера, которое легко рассчитать на основе простой комбинаторной формулы $k = n^m$, где n – мощность множества значений ключа; m – его размер или период.

Рассмотрим взаимосвязь понятий длины и размера ключа на примере шифров перестановки, которые представляют для этого хорошую модель.

3.1. Длина ключа и размер блока

Шифр перестановки является фундаментально блочным в отличие от остальных шифров, блочных технологически. Процедура перестановки невозможна при отсутствии границ. Будем использовать термины «длина» и «размер» для блока как синонимы.

На основании *утверждения 1* мы знаем, что длина блока и размер ключа совпадают (но для ключа в данном случае будем использовать только термин «размер»). Обычно размер n и длину ключа отождествляют, и на основании этого по известной формуле комбинаторики определяют число возможных перестановок:

$$p = n! \tag{3.1}$$

Но это неверно. Если размер ключа равен единице, то очевидно, что шифрования не происходит. Но точно так же не происходит никакого шифрования и при сколь угодно большом n, если ключ имеет такой вид, как показано на рис. 22.

1	2	3	4	5	6	7	8	9	10	11	12	…	n

Рис. 22

Примем соглашение, что *длина* ключа (определяемая в терминах стойкости шифра) в обоих случаях одинакова, и равна единице.

Не совпадают длина ключа и размер блока и в общем случае.

3.2. Длина ключа и шифруемый текст

Результат шифрования всегда зависит и от шифруемого текста. Например, если весь блок шифруемого текста состоит из одного символа, то при любом ключе и его содержании никаких перестановок в тексте мы не увидим (пусть они и происходят технологически). Перестановка одинаковых символов абсолютно бесполезна. В то же время известно, что в текстах на естественных языках существует значительная повторяемость символов. Число перестановок зависит не только от размера ключа, но и от повторяемости символов в тексте, и должно определяться на основе иной формулы комбинаторики, чем (3.1), а именно на основе формулы

$$p = \frac{n!}{i_1!i_2!...i_k!}, \tag{3.2}$$

где i_k – количество повторений символа k.

В соответствии с (3.2) число возможных перестановок, а значит, и длина ключа, зависят как от размеров и содержания ключа, так и от самого шифруемого текста (для различных текстов заданная ключом перестановка будет или не будет работать, или будет работать частично).

Перестановке могут подвергаться как отдельные символы, так и группы символов – число возможных перестановок в зависимости от этого меняется:

{«стол», «стул»} – 2!, {с,т,о,л,у} – 5!

При этом в соответствии с формулой (3.2) оно максимально и равно (n!), если все символы (группы) в тексте различны, снижается при появлении повторяющихся символов (групп), и минимально и равно единице, если повторяется один символ (группа).

Так как любой текст можно представить в виде двоичной последовательности, и именно в таком виде он и обрабатывается сегодня с помощью ЭВМ, приведем следующую оценку возможных перестановок для двоичного сообщения (состоящего из «0» и/или «1») длиной, например, 8:

1) 0 и 1 поровну: $\frac{n!}{(n/2)!(n/2)!} = \frac{8!}{4!4!} = 70;$

2) все 0 или 1: $\frac{n!}{n!} = 1;$

3) есть преобладание: $\frac{8!}{2!6!} = 28.$

Следовательно, для устойчивого шифрования перестановкой необходимо иметь такую двоичную кодировку, в которой одному исходному символу всегда соответствует равное количество «0» и «1». Назовем такую кодировку *защищенной*.

Особенность рассмотренной проблемы заключается в том, что каждый блок текста имеет свой алфавит и свою повторяемость символов, например:

НОВО СОСН ЕНСКИЙ
3 2 1 4 3 2 1 4 …
ВОНО СОСН …

Тогда один и тот же ключ перестановки в пределах шифруемого текста может иметь различную длину.

Определить заранее ее невозможно. Мы можем оценить только вероятность перестановки для каждого *m*-го блока текста в зависимости от размера ключа *n*, например, по формулам

$$p_m^{n,k} = \frac{(n - \sum_{r=j+1}^{k} i_r)!}{i_1!i_2!...i_j!} - 1; \tag{3.3}$$

$$p^0 = n! - 1. \tag{3.4}$$

$$p_m^n = \frac{p_m^{n,k}}{p^0}. \tag{3.5}$$

Здесь *r* – количество символов, не несущих смысловой нагрузки: пробелы, знаки пунктуации, цифры и т. д. Необходимость такой поправки демонстрирует пример с неполностью заполненным (как правило, последним) блоком (рис. 23).

Н				
	Г			
		Т		
			У	

Н..Г..Т.У..

Н				
	Г			
	Т			
		У		

… Н..Г..Т..У..

Рис. 23

«Длину» ключа для данного блока m в соответствии с (3.1) – (3.5) можно приблизительно определить как

$$L_m^n = p_m^n \times n. \tag{3.6}$$

При вычислениях по формулам (3.1)–(3.6) будем полагать, что $n = 1$ во всех случаях, когда *длина* ключа равна 1.

Хотя общей длины ключа перестановки для всех блоков шифруемого текста не существует, можно определить статистику конкретного ключа для фиксированного текста: максимальные и минимальные значения, средние и отклонения от них, другие статистические характеристики, которые и будут представлять общую «длину» ключа, но только для данного конкретного текста. Для другого текста они будут иными.

Можно сделать вывод, что шифр перестановки – это шифр, в котором ключ формируется из открытого текста в соответствии с фиксированным размером и содержанием заданного вектора инициализации (ключа перестановки).

Очевидно, что перестановка очень чувствительна к специально подобранному тексту. Если на вход шифра перестановки подать текст из одного символа (условно «пустой»), то перестановка никак себя не проявит. С другой стороны, при попытке зашифровать перестановкой любую упорядоченную символьную последовательность (например, 0123456 …) на выходе получится ключ.

Подобрать ключ перестановки *идеальной длины* для произвольного текста невозможно. Эффективная длина ключа, или просто эффективность ключа, для разных блоков текста при перестановке различна.

Возникают следующие практические вопросы.

1. Какой размер ключа лучше выбирать для шифра перестановки – длинный или короткий?

2. Что лучше – повторные перестановки или одна, но с «хорошим» ключом?

3. Как влияет на перестановку «связность» ключа, под которой будем понимать перестановку символов в блоке по цепочке (1-й на место 5-го, 5-й на место 7-го и т. д.) и длину такой (таких) цепочки? Слабосвязным будет ключ, в котором все символы переставляются парами, полносвязным, когда все символы переставляются по одной цепочке, длина которой равна размеру ключа.

Однозначного ответа на эти вопросы нет. Сравнение максимально возможного числа перестановок и среднего числа возможных перестановок, подсчитанных по формулам (3.1) и (3.2) для одного и того же

текста (рис. 24), показывает теоретически ожидаемый рост переборной устойчивости ключа вместе с его размером. Реально же происходит сокращение возможного числа ключей относительно теоретического максимума и рост вероятности появления подходящих дубликатов (см. рис. 24).

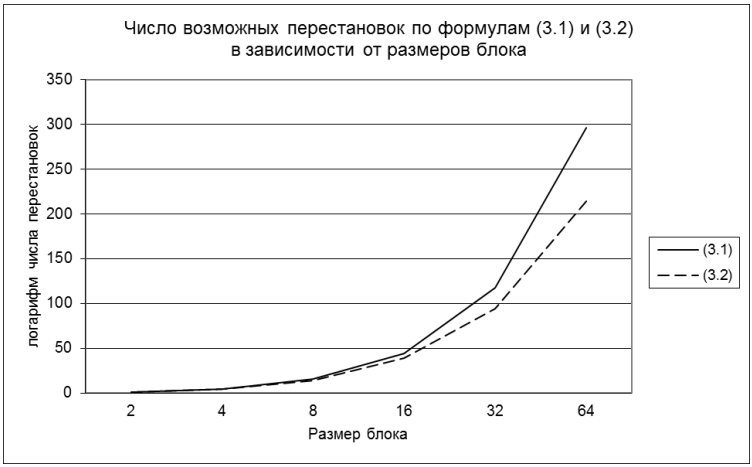

Рис. 24

Повторные перестановки позволяют при определенных условиях быстро нарастить размер реального ключа (блока), используя небольшие длины промежуточных ключей. Но при росте размера блока (рис. 25) увеличиваются возможные вариации эффективности ключа (см. рис. 25, верхний график – блок размером 64, следующий – размером 32 и т. д.), а при наличии неполностью заполненного блока она тем быстрее стремится к нулю, чем больше размер блока. Это поясняет рис. 23.

Что касается «связности» ключа, то из формул комбинаторики следует, что перестановок парами меньше, чем перестановок по всему размеру ключа, т. е. связный ключ лучше, чем слабосвязный, а случайный ключ – ближе к полносвязному.

Важно также помнить, что при повторных перестановках может быть частично или полностью восстановлен исходный текст.

Фундаментальные особенности перестановок, которые следует учитывать при шифровании: блочность; различная эффективность

ключа для разных блоков открытого текста; возможность раскрытия шифра при повторных перестановках.

Рис. 25

Решить проблемы шифра перестановки внутри самого шифра не представляется возможным. Устранить указанные недостатки можно только путем специальной подготовки исходного текста. Для этого в нем нужно убрать повторяемость символов и использовать защищенную двоичную кодировку.

Контрольные вопросы

1. Как определяется стойкость шифра?
2. В каких единицах измеряется длина ключа?
3. В каких единицах измеряется стойкость шифра?
4. Что такое расстояние единственности?
5 В каких единицах измеряется переборная стойкость шифра?
6. В чем различие между длиной и размером ключа?
7. Что такое период ключа?
8. В чем заключаются фундаментальные особенности перестановки?
9. Приведите примеры чувствительности перестановки к специально подобранному тексту.

10. В чем заключается зависимости длины ключа перестановки от шифруемого текста?

11. В чем особенность шифрования перестановкой не полностью заполненного блока текста?

12. Какими способами можно повысить эффективность шифра перестановки?

13. Что такое защищенная кодировка и в чем ее смысл?

14. Как и в каких единицах измеряется переборная стойкость шифра?

Задание на лабораторный практикум

«*Особенности перестановки. Длина ключа*»

1. Зашифровать специально подобранный текст шифром из лабораторного практикума 2, длину текста взять равной трем блокам:

а) текст, состоящий из одного повторяющегося символа (не пробел);

б) текст, состоящий из всех разных символов (без пробелов – цифры, символы русского, английского алфавитов – строчные и заглавные, символы пунктуации и т. д.).

Для пунктов а) и б) рассчитать длину ключа по формулам (3.1) и (3.2) или (3.6).

2. Зашифровать восстановленный текст из лабораторного практикума 1 перестановкой из практикума 2. Сравнить частоты символов, индексы совпадения и частоты биграмм для зашифрованного и закодированного текстов.

3. Для текста из лабораторного практикума 1 рассчитать данные по длине ключа по формулам (3.1) и (3.2) или (3.6) для ключей с размерами: 2, 4, 8, 16, 32, 64.

4. Повторить п. 3 для ключей размером 16, 32, предварительно удалив из текста пробелы.

4. ШИФРЫ ЗАМЕНЫ

Шифрами замены называются шифры, которые заменяют символы открытого текста символами того же или другого алфавита. Возможны замены знаков или групп знаков. Будем рассматривать шифры замены знаков. Их рассмотрение обычно начинается с систем кодирования.

4.1. Системы кодирования Цезаря, Полибия, простой замены

Система кодирования Цезаря основана на замене символа открытого текста символом того же алфавита, но сдвинутым относительно первого на одну позицию (рис. 26).

Исходный алфавит: А Б В…

Замещающий алфавит: А Б В Г … Результат замены АВВА – БГГБ

Рис. 26

Алфавит засматривается как упорядоченный и циклически замкнутый. Возможны модификации системы Цезаря путем изменения величины и направления сдвига, который при этом должен оставаться постоянным.

Система Полибия разработана греческим математиком Полибием (II в. до н.э.). Он предложил заполнять символами алфавита (греческого, 25 букв) матрицу 5×5, причем случайным образом.

Для кодирования символ исходного текста находится по таблице, а затем заменяется символом, находящимся в том же столбце, но на одну строку выше (рис. 27).

Столбцы считаются циклически замкнутыми. При восстановлении исходного текста находится символ закодированного текста по таблице и заменяется на символ, стоящий в том же столбце, но на одну строку ниже. Подход, предложенный в методе Полибия, может использоваться для любых алфавитов и произвольных матриц, а не только для квадратных.

Таблица замены

	…	…	…	
	О	Р	С	
	А	Б	В	
	…	…	…	

Рис. 27

К системам кодирования относится и метод простой замены, в соответствии с которым символы исходного текста заменяются символами того же или другого алфавита, сопоставленными, возможно, случайным образом (рис. 28).

Исходный алфавит: А Б В … Я

Замещающий алфавит: Я Р С … У Результат замены АБВ – ЯРС

Рис. 28

Для всех этих методов (Цезаря, Полибия, простой замены) существует взаимооднозначное соответствие между символом исходного текста и его кодом, т. е. выполняются условия кодирования.

Нетрудно увидеть, что и система Цезаря, и система Полибия являются частными случаями простой замены, и все они для замены используют один алфавит. Шифры, использующие для замены один алфавит, называются одноалфавитными, или моноалфавитными шифрами.

Объяснение использования систем кодирования в роли шифров может быть следующим: как известно, при сокращении объема открытого текста до величины, меньшей расстояния единственности [1], неопределенность (неоднозначность) восстановления текста любыми методами его анализа становится неустранимой. И даже зная, что перед нами закодированный, а не зашифрованный текст, восстановить его *без ключа*, не имея дополнительной информации, мы не сможем.

Отметим, что для восстановления текста, закодированного системой Цезаря, в худшем случае требуется $2n$ переборов (если значение сдвига может принимать значения от 1 до n и менять направление), а текст, закодированный по методу Полибия или простой замены, имеет $n!$ вариантов кодировки.

При достаточном объеме закодированного текста все частотные характеристики исходного сообщения сохраняются, и текст может быть восстановлен логико-вероятностными, вычислительными и/или статистическими методами.

4.2. Шифры Гронсфельда, Виженера, Вернама и гаммирования

Рассмотрение многоалфавитных шифров замены начнем *с шифра Гронсфельда* (Германия, XVII в.). В соответствии с ним для каждого символа текста определяется свой сдвиг, задаваемый десятичной цифрой ключа (рис. 29).

	1 2 3 4 5 6 7
Алфавит:	АБ В ГД ЕЖ …
Открытый текст:	АБВ
Ключ:	2 4 3
Шифротекст:	ВЕЕ

Рис. 29

Алфавит рассматривается как упорядоченный и циклически замкнутый.

При расшифровывании производится обратный сдвиг символа шифротекста по цифре ключа.

Для *выравнивания размера* ключа и открытого текста обычно используется следующее *правило*:

1) если ключ короче шифруемого текста, то он циклически повторяется;

2) если ключ длиннее открытого текста, то неиспользуемая часть ключа отбрасывается.

Например: А Б В Г ДЕЖ
2 4 3 2 4 3 2 <4 3 > – отбрасывается.
[повтор]

В шифре Гронсфельда наглядно представлены два элемента, отличающие настоящий шифр замены от систем кодирования: появление случайного сочетания значения ключа и символа открытого текста, а также использование для замены более одного алфавита (такие шифры называют много- или полиалфавитными).

Шифрование по Гронсфельду можно описать с помощью таблицы замен (рис. 30).

№ п/п	АБВ Г…. …. Я
0 1 2 … 9	АБВ Г…. …. Я БВ Г…. …. ЯА ВГ… …. ЯАБ …. ….

Рис. 30

Символ открытого текста выбирается в первой строке таблицы, фиксируя ее определенный столбец. Цифра ключа фиксирует строку таблицы, на пересечении которой с выбранным ранее столбцом находится заменяющий символ открытого текста символ шифротекста.

Из таблицы, изображенной на рис. 30, хорошо виден и циклический сдвиг алфавитов, и ограниченность ключа шифра Гронсфельда.

Шифр Гронсфельда является частным случаем шифра Виженера (Франция, XVI в.), который также связывают с именами Тритемия (Германия, XVI в.) и Беллазо (Белласо, Италия, XVI в.). Свое постоянное название шифр Виженера имеет с XIX в.

Шифр Виженера использует ключ, значения которого расширены до размеров алфавита. Цифрами ключа являются символы того же алфавита, который используется и для открытого текста, и для шифротекста, т. е. таблица замен является квадратной $n \times n$, где n – число символов алфавита (рис. 31).

Ключ		0 1 2 31 АБВ …. …. Я
0	А	АБВ …. …. Я
1	Б	БВ …. …. ЯА
…	…	…
31	Я	ЯАБВ ….

Рис. 31

Сдвиг алфавитов по-прежнему циклический и на один символ, как и в шифре Гронсфельда.

Отметим две важных особенности метода Виженера.

1. Код шифротекста может быть получен по формуле

$$z = (p + k) \bmod N, \tag{4.1}$$

где p, k –коды символов открытого текста и ключа соответственно; N – мощность алфавита.

Расшифровка может быть выполнена по формулам (рис. 32)

$$p = \begin{cases} N + z - k, & z < k, \\ z - k, & z \geq k. \end{cases} \tag{4.2}$$

0 12	$N = 3$	Тогда:	
0 АБВ		АБВ – 012	211
1 БВА		ВАВ – 202	202
2 ВАБ		ВББ = 211	012 (3 + 1 – 2)
		ВББ	АБВ

Рис. 32

2. При использовании в качестве ключа открытого текста, или текста, закодированного в системе Цезаря, получаем шифротекст из одного единственного символа (диагонали таблицы замен, см. рис. 31).

И хотя такой шифротекст абсолютно не раскрываем, применение подобного ключа делает саму систему секретной связи с помощью шифров бессмысленной, так как текст-ключ уже передан получателю по закрытому каналу связи. Шифрование в таком случае *вырождается* в тайнопись.

Однако эти особенности не сохраняются для других более сложных замен, когда сдвиг алфавитов носит не циклический, а иной (например, случайный) характер.

Например, для монофонической замены, когда наиболее встречающиеся в тексте символы заменяются на менее встречающиеся. При использовании данного шифра для одного и того же символа могут использоваться несколько заменяющих символов (рис. 33).

А … О …
Ф … Л …
* … В …
Л … / …

Рис. 33

После шифрования каждого символа соответствующий ему столбец алфавита циклически сдвигается вверх на одну позицию и для каждого символа *по мере его встречаемости в открытом тексте* формируется своя замена.

Монофоническая замена относится к специальному виду *шифров с «автоключом»*. Ими называются шифры, в которых само сообщение или шифротекст используются в качестве ключа. Описать такие шифры только с помощью формул (4.1), (4.2) в общем случае нельзя, так как значения ключа до появления шифруемого теста остаются неопределенными.

Шифры Бофора (Англия, XIX в.) и Вернама (США, XX в.) являются модификациями шифра Виженера.

Шифры Бофора использует следующие формулы вместо (4.1):

$$z = (k - p) \bmod N, \tag{4.3}$$

$$z = (p - k) \bmod N. \tag{4.4}$$

Шифр Вернама – это шифр Виженера, в котором используется случайный «бесконечный» ключ, нормированный по *mod N*.

Свое развитие и современное воплощение метод Виженера – Вернама получил в методе гаммирования, идеально подходящем для компьютерной обработки.

Если в качестве кодов текста и ключа взять их двоичные коды, то и суммирование, и вычитание для операций (4.2) можно заменить одной операцией сложения по модулю два:

$$\begin{aligned} z^{(2)} &= p^{(2)} \oplus k^{(2)}, \\ p^{(2)} &= z^{(2)} \oplus k^{(2)}. \end{aligned} \tag{4.5}$$

Получаем однородный (нет операции вычитания) симметричный простой алгоритм, где ключ называется гаммой, по названию греческой буквы, которая дала название всему *шифру гаммирования* (рис. 34).

Шифрование		Расшифрование	
Открытый текст:	0110011	Шифротекст:	1011110
Ключ (γ):	(+) 1101101	Ключ (γ):	(+) 1101101
Шифротекст:	1011110	Открытый текст:	0110011

Рис. 34

4.3. Взаимосвязь шифров замены

Рассмотренные в разделе 4.1 шифры замены можно представить с помощью таблиц замен (рис. 35).

Как видно из таблиц, изображенных на рис. 35, системы Цезаря и Полибия эквивалентны простой замене, а система Цезаря и шифр Гронсфельда являются частными случаями шифра Виженера.

Шифры Цезаря, Гронсфельда, Виженера и различные их модификации, включая методы Бофора, Вернама и гаммирования, образуют

частный класс шифров замены, которые могут быть реализованы без использования таблиц замен, а только на основе формул (4.1) и (4.2), которые дают общую математическую модель этих шифров.

Таблица замен системы Цезаря

Ключ	Символы алфавита				
	0	1	2		30
0	а	б	в	…	я
1	б	в	г		а

Таблица замен системы Полибия

Ключ	Символы алфавита				
	0	1	2		30
0	а	б	в	…	я
1	о	р	с		в

Таблица простой замены

Ключ	Символы алфавита				
	0	1	2		30
0	а	б	в	…	я
1	я	р	с		у

Рис. 35

Модель шифров (4.1) и (4.2) основана на мощности используемых при шифровании алфавитов и их упорядоченности. Кодами символа p открытого текста и символа z шифротекста являются их порядковые номера в одном алфавите **A** мощности N. Изменяя варианты упорядоченности символов алфавита (т.е. выбирая другой алфавит, но с тем же множеством знаков), мы можем влиять на значения p, z. Меняя значения ключа k, который задает сдвиг по алфавиту, можно получать различные системы кодирования и шифрования.

Так, для системы Цезаря достаточно положить k = const (0, 1, 2, …n). При $k = 0$ шифрования не происходит. Шифр Гронсфельда получаем, если значения k принадлежат {0, 1, 2,…, 9}, шифр Виженера – если значения k принадлежат {0, 1, 2,…, N–1}. В любом случае при использовании формул (4.1) и (4.2) по умолчанию считается, что для

шифра *определена* таблица замен типа Виженера циклически сдвинутых алфавитов, при этом не надо забывать о направлении и размере сдвига.

Для того чтобы изменить направление сдвига в рассмотренных и других случаях, достаточно поменять порядок применения операций (4.1) и (4.2) или, что еще проще, допустить отрицательные значения k (математически это эквивалентно, но последний вариант проще в реализации).

Ограничения по использованию модели (4.1) и (4.2) определяются применяемыми алфавитами и способами формирования и применения ключа.

4.4. Алфавиты и композиция замен

Алфавитом называется множество упорядоченных неповторяющихся знаков (символов, букв). Могут рассматриваться алфавит языка $\mathbf{A}_L$ или сообщения (текста) $\mathbf{T} - \mathbf{A}_T$.

Значением знака (символа, буквы) является его порядковый номер в алфавите (код знака).

Возможна и известна из практики множественность знаков одного значения. Тогда наиболее общее определение алфавита следующее.

Алфавитом называется множество упорядоченных множеств неповторяющихся знаков (символов, букв), имеющих одинаковое значение.

Из этого определения следует, что каждый символ (буква) алфавита языка может быть представлен в тексте сообщения **T** не одним, а множеством знаков, что хорошо известно из практики.

В текстах могут использоваться все знаки букв. Это приводит к «раздуванию» (знаковой, графической *инфляции*) исходного алфавита языка и неопределенности связи между мощностью алфавита языка $\mathbf{A}_L$ и мощностью алфавита сообщения $\mathbf{A}_T$, если последний основывается на числе используемых в тексте знаков.

Инфляция исходного алфавита существенно осложняет криптоанализ текстов, и может рассматриваться как полезная подготовка исходного текста перед шифрованием.

Из определения следует, что алфавит – это бинарное отношение между множеством последовательных натуральных чисел $\mathbf{N} = \{1, 2, \ldots, n\}$ и множеством знаков $\mathbf{C} = \{\mathbf{C}^i\}$, $i = 1...n$, $\mathbf{A} = \mathbf{N} \times \mathbf{C}$, $\mathbf{A}=\{<i, \mathbf{C}^i>\}$,

$\mathbf{C}^i = \{c_1{}^i, c_2{}^i, \ldots, c_m^i\}$, $\bigcap_{i=1}$, где c_m^i – одиночный знак (символ, буква). Более того, **A** – это отношение упорядочения и его математическое значение – это *ось координат* **A** на множестве знаков **C**. Мощность алфавита **A** равна |**A**| = |**N**| = *n*.

Количества знаков одного значения (мощность $\mathbf{C}^i$) могут быть различными. Для упрощения уравняем мощности всех $\mathbf{C}^i$, разрешив произвольное дополнение их уже существующими знаками до максимального значения *m*. Выберем из каждого $\mathbf{C}^i$ один знак $a_i = c_j^i$ и образуем из них алфавит $\mathbf{A}^S$, который будем считать *стандартным*. Будем называть $\mathbf{C}^i$ вектором знаков символа (буквы) или просто вектором символа (буквы), если на нем задано некоторое упорядочение такое, что первый символ в этом упорядочении является символом стандартного алфавита $\mathbf{A}^S$.

Будем под *символом* алфавита c_1^i понимать любой знак из множества $\mathbf{C}^i$, под *буквой* алфавита – стандартный знак буквы и/или ее значение – *код*, под *значением буквы* (*символа*) – только его порядковый номер (код) в алфавите.

Под отождествлением символов будем понимать отношение двух знаков такое, что первый знак пары отождествляет второй, т. е. приписывает ему значение первого. Под совпадением символов понимают использование одинаковых знаков, под равенством символов – равенство их значений в одном алфавите вне зависимости от используемых знаков.

Все символы $\mathbf{C}^i$ равны и тождественны стандартному символу – букве алфавита.

Будем называть *входным* (*исходным*) *алфавитом* алфавит текстового сообщения, поступающего на вход алгоритма шифрования, *выходным алфавитом* – алфавит сформированного шифротекста. Будем считать, что входной (исходный) алфавит является стандартным.

Тогда шифрование заменой можно рассматривать как преобразование над композицией входного алфавита $\mathbf{A}_T^S = \mathbf{A}_0$ и алфавитов ключа $\mathbf{A}_i$, i = 1, 2,…,m, образующих выходной алфавит $\mathbf{A}_c$, и определить следующую таблицу композиционной замены (рис. 36).

A	**K**	Значения символов алфавита									
		0	1	2	3		15	16		27	$N = 30$
A_0	0	а	б	в	г		п	р		ы	я
A_1	1	α_1	β_1	γ_1	δ_1		п	σ_1		ы	λ_1
A_2	2	α_2	β_2	γ_2	δ_2		п	σ_2		ы	λ_2
A_3	3	α_3	β_3	γ_3	δ_3		п	σ_3		ы	λ_3
		…	…	…	…		…	…		…	…
A_k	k	α_k	β_k	γ_k	δ_k		п	σ_k		ы	λ_k
A_{k+1}	k+1	α_1	β_{k+1}	γ_{k+1}	δ_{k+1}		п	σ_{k+1}		ы	λ_{k+1}
A_{k+2}	k+2	α_2	β_{k+2}	γ_1	δ_{k+2}		п	σ_{k+2}		ы	λ_{k+2}
A_{k+3}	k+3	α_3	β_{k+3}	γ_2	δ_1		п	σ_{k+3}		ы	λ_{k+3}
A_{k+4}	k+4	α_4	β_{k+4}	γ_3	δ_2		п	σ_{k+4}		ы	λ_{k+4}
A_m	m	α_1	β_m	γ_7	δ_3		п	σ_m		ы	λ_m
$\mathbf{A}_c = \bigcup$											
		C^0	C^1	C^2	C^3		C^{15}	C^{16}		C^{27}	C^{30}

Рис. 36

Столбец **A** в таблице, изображенной на рис. 36, обозначает множество алфавитов, столбец **K** – множество значений ключа. Выделенные в таблице (рис. 36) ячейки показывают мощность множеств $\mathbf{C}^i$, если рассматривать их как *имитацию* алфавитов *букв*, т. е. не содержащую повторов упорядоченную последовательность знаков, имеющих одинаковое значение. Затем по обычному правилу они дополнены до максимального значения m *циклическим расширением*. Но это вовсе не обязательно, и дополнение до m может быть сделано произвольным выбором знаков из $\mathbf{C}^i$.

Выходной алфавит шифра образуется объединением множеств $\mathbf{C}^i$ без символов входного алфавита $\mathbf{A}_0$, на котором (объединении) задано некоторое упорядочение:

$$\mathbf{A}_c = \bigcup_{i=0}^{N} \quad .$$

Шифрование и расшифрование с помощью таблицы, представленной на рис. 36, осуществляется следующим образом.

1. Значения ключа k выбираются случайно из множества натуральных чисел **N** и нормируются по mod(m+1).

2. При шифровании значение символа исходного текста из строки алфавита $\mathbf{A}_0$ определяет столбец таблицы, в котором находится замещающий символ. Значение ключа k определяет строку алфавита $\mathbf{A}_k$, в которой находится замещающий символ. На пересечении данных столбца и строки выбирается замещающий исходный символ алфавита $\mathbf{A}_0$ символ алфавита $\mathbf{A}_k$.

3. При расшифровании в строке $\mathbf{A}_k$, определяемой значением ключа k, находится символ шифротекста, к которому применяется данный ключ. Значение найденного символа шифротекста $\mathbf{A}_k$ определяет столбец (значение) символа исходного алфавита $\mathbf{A}_0$. Этот символ заменяет символ шифротекста из алфавита $\mathbf{A}_k$.

Назовем такое шифрование пп. 1–3 композиционной заменой. Для его обратимости необходимо выполнение следующего условия.

Утверждение 2. Для обратимого шифрования композиционной заменой необходимо и достаточно, чтобы в каждой строке ключа не было совпадающих знаков.

Другими словами, необходимо, чтобы каждая строка таблицы являлась алфавитом. Совпадение знаков в столбцах допускается, если это не противоречит условию *утверждения 2*. В частности, на рис. 36 показано, что символы (п, ы) входного алфавита $\mathbf{A_0}$ не заменяются ни при каких значениях ключа, и переносятся в шифротекст без изменения.

Из *утверждения 2* следует, что таблица композиционных замен (рис. 36) образуется из алфавитов полных или частичных простых замен в исходном и/или другом алфавите. Максимально возможный теоретический размер таблицы определяется не только числом таких алфавитов, но также различными их сочетаниями. В практических целях можно считать данный размер неограниченным.

Шифры замены, в которых входной и выходной алфавиты равны, называются моноалфавитными, или одноалфавитными. Они могут быть реализованы на основе модели (4.1) и (4.2). Равенство алфавитов предполагает, что где бы в таблице замен не находился один и тот же символ, он будет иметь одно и то же значение, получаемое по формулам (4.1) и (4.2). Для этого алфавиты таблицы замен должны быть связаны с помощью одинакового циклического сдвига или другим регулярным образом.

По модели (4.1) и (4.2) могут быть реализованы и шифры, у которых входной и выходной алфавиты полностью различаются наборами знаков и даже их количеством. Для этого оба алфавита должны быть формально объединены в один алфавит и введена поправка для ключа, равная мощности входного алфавита. Существуют и другие способы расширения области применения модели (4.1) и (4.2), которые могут быть подобраны индивидуально в некоторых частных случаях.

Композиционная замена – это многоалфавитная замена, которая в общем случае не может быть описана уравнениями (4.1) и (4.2), так как в ней отсутствует взаимосвязь алфавитов ключа и допускается появление в таблице замен совпадающих знаков, имеющих разное значение. Шифры (4.1) и (4.2) являются частными случаями композиционной замены.

Утверждение 3. Все симметричные шифры знаковой замены функционально эквивалентны композиционному шифру замены.

Справедливость этого утверждения уже была показана для шифров из разделов 4.1–4.3 и модели (4.1) и (4.2). Эти шифры не только функционально, но и строго эквиваленты композиционному шифру, так как используют таблицы замен, являющиеся частными случаями композиционной таблицы, изображенной на рис. 36.

Покажем применимость композиционной замены в случае монофонической замены. Монофоническая замена – это специальный шифр с «автоключом». Она представляет собой инфляционное преобразование алфавита исходного сообщения в более мощный алфавит шифротекста. Для ее моделирования необходимо в каждом столбце композиционной таблицы замен (см. рис. 36) иметь несовпадающие друг с другом знаки. Но этого недостаточно, так как значения случайного ключа могут повториться для одного и того же исходного символа, и в таком случае возможна одинаковая замена его повторов.

Для того чтобы точно заменить монофоническую замену композиционной, можно использовать два подхода, связанные с подготовкой ключа.

1. Для каждого шифруемого текста подготовить ключ монофонической замены на основе таблицы замен. Передать ключ принимающей стороне, и только затем отправлять зашифрованное сообщение.

2. В таблице замен использовать только неповторяющиеся символы. Тогда ключом будет сама таблица замен (а не значения k), а восстановление текста должно осуществляться по принадлежности того или иного символа множеству $\mathbf{C}^i$.

И в том и другом случае такая замена функционально эквивалентна композиционной замене.

Другой специальный шифр замены с «автоключом» – это, как ни странно, шифр перестановки. Легко показать, что шифр перестановки – это шифр замены. Так как коды символов *p, z* до и после перестановки известны, то

$$z = \big(p + [z - p]\big) \bmod N = (p + k) \bmod N,$$

$$p = \big(z - [z - p]\big) \bmod N = (z - k) \bmod N.$$

Вычитая коды открытого текста из кодов полученного перестановкой шифротекста, мы получим ключ замены, функционально эквивалентной данному шифру перестановки на данном открытом тексте. Тогда для перестановки допустим следующий вариант: зашифровать перестановкой, расшифровать заменой.

Отсюда вытекает справедливость следующего утверждения.

Утверждение 4. Шифр перестановки является частным случаем шифра замены.

Еще одним распространенным специальным шифром с «автоключом» является шифр гаммирования с обратной связью (рис. 37).

Сравнение двух последних шифров показывает, что перестановка похожа на гаммирование с обратной связью тем, что в обоих случаях на основе некоторого общего стандартного ключа, который можно рассматривать как *вектор инициализации*, ключ шифра формируется из открытого или закрытого текста.

Открытый текст:		T_1	T_2	T_3		T_n
Ключ:	(+)	$\gamma_1 = \gamma_0$	$\gamma_2 = C_1$	$\gamma_3 = C_2$	…	$\gamma_n = C_{n-1}$
Шифротекст:		C_1	C_2	C_3		C_n

Рис. 37

Применение повторной замены для ряда шифров рассмотрел Шеннон. Для шифра Виженера он показал, что его последовательное применение с разными размерами ключа (периодами) эквивалентно суммированию значений этих ключей на размере общего ключа, равном НОК использованных ключей. Для простой замены и последующего шифрования по методу Виженера достаточно образовать таблицу замен шифра Виженера по измененному простой заменой алфавиту [1].

Из рассмотренных *утверждений 1–4* следует сделать вывод, что базовым криптографическим преобразованием является замена знаков, в основе которой лежит простая замена. В зависимости от способа подготовки ключа шифры замены подразделяются на замены со случайным ключом и «автоключом». Для первых следует выделить шифры модели (4.1) и (4.2) и особенно шифр гаммирования, как наиболее подходящий для компьютерного применения. Специальные шифры с «автоключом» подразделяются на шифры, в которых ключ формируется на основе открытого текста – монофоническая замена и шифр перестановки. А также шифры с «автоключом», в которых ключ формируется на основе и открытого текста, и шифротекста, – это шифры с обратной связью, и особенно – шифр гаммирования с обратной связью. Другие «классические» шифры в большей степени направлены на алгоритмическую подготовку ключа.

Контрольные вопросы

1. Какие шифры называются шифрами замены?
2. Опишите шифр:
   а) Цезаря;
   б) Полибия;
   в) простой замены;
   г) Гронсфельда;
   д) Виженера.
3. Чем системы Цезаря, Полибия и простой замены отличаются от других шифров замены?
4. Приведите формулу общей математической модели шифров Цезаря, Гронсфелда, Виженера.
5. Какие шифры называются шифрами Бофора?
6. Какой шифр называется шифром Вернама?
7. Постройте собственный пример шифрования и расшифрования по формулам (4.2), отличающийся от приведенного в настоящее пособии?
8. Какой шифр называется шифром гаммирования, с какими известными шифрами и как он связан?
9. Что называется инфляцией алфавита?
10. В чем основная особенность шифра гаммирования?
11. Какое условие должно выполняться для обратимого шифрования композицией замен?

12. Какой алфавит называется входным (исходным) для шифра, а какой – выходным алфавитом?

13. Какие шифры называются одноалфавитными, моноалфавитными, полиалфавитными, многоалфавитными?

14. Как получить ключ шифра замены при шифровании перестановкой?

15. Частным случаем каких шифров является шифр перестановки?

16. Какой шифр называется шифром гаммирования с обратной связью?

17. Что общего у шифров монофонической замены, перестановки и гаммирования с обратной связью?

18. Как объединяются повторные применения шифров:

а) Цезаря;

б) Виженера;

в) простой замены и Виженера?

Задание на лабораторный практикум

«*Шифры замены*»

1. Закодировать исходный текст в системах кодирования Цезаря, Полибия и простой замены. Получить частотные характеристики повторяемости символов и биграмм исходного и закодированных текстов и отобразить их на графике Excel.

2. Зашифровать исходный текст шифрами Гронсфельда и Виженера.

3. Многоалфавитная замена. Взять первые 50 символов исходного текста. Расширить русский алфавит буквами английского алфавита с добавлением десятичных цифр. Сформировать для выделенного фрагмента текста ключ замены, максимально использующий оба алфавита. Зашифровать текст.

4. Монофоническая замена. Сформировать для выделенного фрагмента текста ключ замены, дающий одинаковую частоту встречаемости символов. Зашифровать и расшифровать текст.

5. Связь шифров перестановки и замены. Зашифровать перестановкой исходный текст и подготовить «вычитанием» исходного и шифротекста ключ замены. Используя полученный ключ, расшифровать исходную перестановку алгоритмом замены. Сравнить результаты.

5. СВОЙСТВА ШИФРОВ

Рассмотрим следующие определения, введенные Шенноном на основе его математической модели шифров и информационной меры – *энтропии H* [1].

Шифр называется *совершенно секретным*, или *совершенным*, если апостериорное распределение вероятностей исходного случайного сообщения совпадает с априорным распределением вероятностей.

Следовательно, шифротекст не добавляет никакой информации о переданном сообщении.

Необходимым и достаточным условием совершенства шифра является независимость условного распределения вероятностей шифротекста для некоторого сообщения от вероятности самого сообщения. При этом количество информации по Шеннону, содержащееся в шифротексте Y об исходном тексте X, равно нулю.

Следовательно, шифры перестановки, монофонической замены и гаммирования с обратной связью не могут быть совершенными шифрами по определению, так как зависят от открытого текста.

Для получения совершенного шифра необходимо, чтобы выполнялись следующие неравенства энтропий ключа, сообщения X и шифротекста Y [1]:

$$H(k) \geq H(X),$$

$$H(k) \geq H(Y).$$

Для выполнения этих неравенств необходимо, чтобы длина ключа m была не меньше длины шифруемого текста n: $m \geq n$, или в минимальном случае $m = n$.

Для конечных сообщений и ключей необходимым и достаточным условием получения совершенного шифра является равновероятность используемых ключей. Шифр Вернама является совершенным шифром по Шеннону, следовательно, не дает никакой информации об исходном сообщении.

Шифр называется *идеально стойким*, если невозможно однозначно определить открытый текст при известном шифротексте сколь угодно большой длины.

Совершенный шифр является идеально стойким.

Для конечных сообщений существует возможность неуменьшения стойкости шифра с ростом длины сообщения. Шифры, реализующие

это свойство, называются *абсолютно стойкими*. Такое свойство достигается при выполнении следующих условий:

1) полной случайностью ключа;

2) равенством размера ключа размеру текста;

3) однократным использованием ключа.

Обобщая эти понятия, Шеннон ввел понятие расстояния единственности шифров, и на его основе определил, что стойкость шифра – это тот минимальный объем зашифрованного текста, статистическим анализом которого можно вскрыть исходный текст.

При объеме текста, меньшем этого расстояния, невозможно получить однозначное решение шифра, а только неоднозначные *правильные* варианты.

Шеннон попытался получить общий способ вычисления такой стойкости шифра. Однако ввиду сложности задачи этого сделать не удалось, и были получены только оценки расстояния единственности для некоторых шифров.

По Шеннону, длина ключа в терминах стойкости шифра (или расстояния единственности) для шифра Виженера составляет $2d$, где d – период или размер ключа. Переборная стойкость ключа такого размера для англоязычных текстов составит 26^d. Разница в оценках очевидна. Ценность понятия длины ключа заключается в том, что она точно указывает объем текста, который на данном ключе можно шифровать – например, если $d = 10$, то такой объем составит всего 20 символов. А число 26^{10} в данном случае дает информацию только о большом возможном количестве ключей, и, скорее, вводит в заблуждение, чем говорит о стойкости шифрования.

Для шифра Цезаря, который является частным случаем шифра Виженера с периодом $d = 1$, длина ключа составит $2 \times 1 = 2$, а переборная стойкость – $26^1 = 26$.

Для простой замены (d=1) имеем оценку длины в $53d = 53 \times 1 = 53$ символа. Для простой замены символов русскоязычных текстов в [4] приведена оценка в 48 символов. Переборная стойкость простой замены для англоязычных текстов равна $(26!)^1 = 10^{26,3}$, для русскоязычных текстов – (32!).

Таким образом, стойкость шифра Цезаря ограничена двухбуквенным сочетанием. Для его вскрытия достаточно идентифицировать всего одну биграмму языка. И это очевидно, так как все биграммы состоят из символов, имеющих разные порядковые номера в исходном алфавите, а по шифру Цезаря эти символы должны иметь одинаковый сдвиг.

Стойкость (расстояние единственности) системы простой замены составляет примерно 50 символов текста.

Понятие совершенной и иной стойкости шифра не должно вводить в заблуждение. Оно связано в первую очередь с балансом всех трех элементов шифрования: открытого текста, шифра и ключа. В то же время алгоритмы шифрования имеют свои собственные недостатки.

Для шифров замены рассмотрим особенности алгоритмов шифрования на примере шифра гаммирования.

Основой алгоритма гаммирования является операция побитового сложения по модулю 2. Двоичный код исходного символа складывается с двоичным кодом «гаммы» и получается двоичный код символа шифротекста (рис. 38, *а*).

101010	001101	001101	000000	101010
«+» 100111	«+» 100111	«+» 101010	«+» 100111	«+» 000000
001101	101010	100111	100111	101010
а	*б*	*в*	*г*	*д*

Рис. 38

Если сложить «шифротекст» и «открытый текст» (рис. 38, *в*), результатом будет не что иное, как «ключ». Именно отсюда возникает одно из основных правил практической криптографии – никогда не накладывать шифр на стандартную информацию, ибо наличие открытого текста (в данном случае – стандартной общеизвестной информации) и полученного по нему шифра в руках пользователей, не имеющих ключа, полностью разрушает шифр.

Происхождение другого практического правила – требования отключения аппаратуры (программы) шифрования при отсутствии шифруемого текста иллюстрирует пример (рис. 38, *г*). Если ваша аппаратура не будет выключена (или программа будет шифровать «все подряд»), в открытый канал связи в качестве шифротекста будет передан ключ. Последствия очевидны.

Длина ключа, который сам является информационной единицей, измеряется в тех же единицах измерения, что и открытая информация, т. е. в битах. Таким образом, длина «нулевого» ключа (рис. 38, *д*) равна одному биту (который циклически повторяется). Такую же длину имеет и «единичный» двоичный ключ (рис. 39). Нетрудно увидеть, что коды символов шифротекта при «нулевом» ключе полностью совпа-

дают с кодами символов открытого текста, но также и при «единичном» ключе *каждому исходному символу будет соответствовать один и тот же символ шифротекста*. Таким образом, ***шифр вырождается в систему кодирования,*** основным отличительным признаком которой и является взаимооднозначное соответствие между объектом кодирования и его кодом (сколько раз повторяется исходный объект, столько раз повторится и его код). В случае «нулевого» ключа имеем вырождение уже и системы кодирования в систему копирования.

	Объект1		Объект2	...	**Объект1**
Код	101010		001101	...	101010
Ключ	«+» 111111	«+»	111111	...	«+» 111111
	010101		110010	...	010101

Рис. 39

Следует обратить внимание, что при любой длине ключа вычислительная трудоемкость алгоритма одинакова. Это означает, что результат шифрования зависит не от алгоритма, а от человека, его применяющего. Механическое шифрование в прикладной криптографии недопустимо.

Другое, имеющее важнейшее прикладное значение, использование операции сложения по модулю 2 связано с так называемым гаммированием с обратной связью (см. рис. 37). В этом случае для каждого последующего блока гаммы (т. е. ключа), кроме первого, используется полученный на предыдущем шаге блок шифротекста. Первоначально схема была предложена для решения проблемы «коротких» ключей. Действительно, взяв небольшой ключ для первого блока, мы затем из него и открытого текста получаем относительно «сильный» ключ (считая, что шифруемые тексты достаточно «случайны») с длиной, фактически равной длине открытого текста. Даже для повторяющейся последовательности исходных символов мы получаем последовательность неповторяющихся символов шифротекста.

Но все-таки для практических приложений главной находкой в данной схеме оказалась совсем не длина ключа. Если внимательно проанализировать рис. 37, то видно, что каждый бит второго блока функционально зависит от первого блока гаммы и *каждого бита* первых двух блоков открытого текста, т. е. *каждый бит последнего блока*

зависит от всех битов открытого текста. А это уже позволяет проанализировать целостность исходного текста вплоть до изменения хотя бы одного бита. Поэтому данная схема получила распространение во всех современных алгоритмах шифрования и хэширования (для целей защиты информации) наряду с базовым алгоритмом.

Однако у базового гаммирования с обратной связью (см. рис. 37) легко увидеть фундаментальный недостаток. В поле ключа («гаммы») располагается результат работы, т. е. шифротекст, но со сдвигом на один блок. Тогда нам достаточно взять шифротекст, сдвинуть его на один блок (причем неважно, в какую сторону), скопировать его в поле ключа и затем сложить в обычном режиме (см. рис. 34) с шифротекстом, чтобы получить открытый текст (за исключением первого блока). Поэтому во всех стандартных алгоритмах шифрования при режимах с обратной связью блок шифротекста предварительно дополнительно шифруется перед использованием в качестве блока гаммы.

Часто используемая схема перестановки в сочетании с гаммированием, которая называется сетью Фейстейля [16], без дополнительных преобразований на третьем шаге дает исходный текст (рис. 40).

L_1	R_1
$\oplus$	
$L_2 = L_1(+)R_1$	$R_2 = L_1$
$\oplus$	
$L_3 = L_2(+)R_2 = L_1(+)R_1(+)L_1 = R_1$	$R_3 = L_2 = L_1(+)R_1$
$\oplus$	
$\boldsymbol{L_4} = L_3 (+)\ R_3 = R_1(+)L_1(+)R_1 = \boldsymbol{L_1}$	$\boldsymbol{R_4} = L_3 = \boldsymbol{R_1}$

Рис. 40

Шифрование бывает блочным или потоковым. Шифрование, но не шифры. Собственно действительно блочным шифром является только шифр перестановки.

Если размер блока шифра равен единице (одному знаку), то получаем потоковое шифрование. При этом знаками могут быть биты, байты, символы или в редких случаях более крупные единицы текста.

Следует помнить, что при таком размере блока некоторые шифры могут вырождаться.

При блочном шифровании размер блока больше единицы. Преобразование осуществляется для всех символов, входящих в блок.

Контрольные вопросы

1. Какой шифр называется совершенно секретным?
2. Какой шифр называется идеально стойким?
3. Какой шифр называется абсолютно стойким?
4. Какова стойкость шифра Виженера?
5. Какова стойкость систем Цезаря и простой замены?
6. На примере шифра гаммирования объясните, почему нельзя накладывать шифр на стандартную информацию?
7. На примере шифра гаммирования поясните, почему нельзя шифровать любой предлагаемый текст?
8. В чем заключается основной недостаток шифра гаммирования с обратной связью?
9. Приведите пример вскрытия исходного текста при шифровании повторными перестановками?
10. Какое шифрование называется блочным, а какое – потоковым?
11. В каких случаях при блочном шифровании шифр может вырождаться?
12. При каком объеме открытого текста система Цезаря и ее производные становятся шифрами?
13. При каком объеме открытого текста система простой замены становится шифром?
14. При каком объеме открытого текста преобразование Виженера является шифром?

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Шеннон К.* Теория связи в секретных системах / К. Шеннон // Работы по теории информации и кибернетике. – М.: ИЛ, 1963. – С. 333–369.

2. *Бабенко Л.К., Ищукова Е.А.* Анализ симметричных криптосистем // Известия ЮФУ. Технические науки. – 2012. – № 12. Т. 137. – С. 136–147.

3. *Бабаш А.В.* Криптографические методы и средства информационной безопасности / А.В. Бабаш, Е.К.Баранова. – М.: РГСУ, 2010.

4. *Борисенко Н.П.* Криптографические методы защиты информации / Н.П. Борисенко, А. В. Гусаров, В. И. Милашенко, С. В. Можин. – Орел: Академия ФСО России, 2007.

5. *Гультяева Т.А.* Основы теории информации и криптографии / Т.А. Гультяева. – Новосибирск: Изд-во НГТУ, 2010.

6. *Жданов О.Н.* Криптоанализ классических шифров / О.Н. Жданов, И.А. Куденкова. – Красноярск: Изд-во Сиб. гос. аэрокосм. ун-та им. акад. М.Ф. Решетнева, 2008.

7. *Зубов А.Ю.* Криптографические методы защиты информации. Совершенные шифры. – М.: Гелиос АРВ, 2005.

8. *Котов Ю.А.* Детерминированная идентификация буквенных биграмм в русскоязычных текстах // Труды СПИИРАН. – 2016. – № 1. – С. 181–197.

9. *Минин И.В.* Криптографические методы защиты информации / И.В. Минин. – Новосибирск: Изд-во НГТУ, 2009.

10. *Мао В.* Современная криптография: теория и практика. – М.: Вильямс, 2005.

11. *Минеев М.П., Чубариков В.Н.* Лекции по арифметическим вопросам криптографии. – М.: Изд-во «Попечительский совет Механико-математического факультета МГУ им. М. В. Ломоносова», 2010.

12. *Осипян В. О.* Криптография в задачах и упражнениях. – М.: Гелиос АРВ, 2004.

13. *Рябко Б. Я., Фионов А.Н.* Криптографические методы защиты информации. – М.: Горячая линия-Телеком, 2010.

14. *Росошек С. К.* Специальные главы математики (Математические основы криптографии). Ч. 2. – Томск: Изд-во ТУСУР, 2005.

15. *Смарт Н.* Криптография. – М.: Техносфера, 2006.

16. *Токарева Н.Н.* Симметричная криптография. Краткий курс. – Новосибирск: Изд-во Новосиб. гос. ун-та, 2012.

ОГЛАВЛЕНИЕ

Котов Юрий Алексеевич

КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ

ШИФРЫ

Учебное пособие

Редактор *Л.Н. Ветчакова*
Выпускающий редактор *И.П. Брованова*
Дизайн обложки *А.В. Ладыжская*
Компьютерная верстка *Н.В. Гаврилова*

Налоговая льгота – Общероссийский классификатор продукции
Издание соответствует коду 95 3000 ОК 005-93 (ОКП)

Подписано в печать 22.06.2016. Формат 60 × 84 1/16. Бумага офсетная
Тираж 50 экз. Уч.-изд. л. 3,48. Печ. л. 3,75. Изд. 72. Заказ № 953
Цена договорная

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20