Оглавление

ВВЕДЕНИЕ

Курс «Криптографические методы защиты информации» является базовым при подготовке специалистов по защите информации. На основе знаний криптографии выстраивается система подготовки специалистов. При этом изучение методов защиты неразрывно связано с изучением возможных атак на алгоритмы и на их реализации. Хорошо известно, что для усвоения материала необходима активная самостоятельная работа студентов. Поэтому представляется целесообразным проведение лабораторных работ по криптоанализу. Работы по анализу таких шифров, как DES, ГОСТ 28147-89, IDEA требуют большого ресурса и для начинающего являются чрезвычайно сложными. В то же время на примерах классических шифров можно проиллюстрировать некоторые важные приемы и методы криптоанализа. Как показывает практика работы, студенты после анализа шифров перестановки, простой замены и Виженера уверенно и достаточно быстро входят в круг идей современной криптографии. Таким образом, настоящее пособие выполняет пропедевтическую функцию. После анализа классических шифров учащиеся успешно изучают современные блочные алгоритмы шифрования, им становятся доступными идеи линейного и дифференциального криптоанализа.

Авторы сочли необходимым теоретические сведения дополнить подробно изложенными примерами выполнения заданий. После изучения теории и ознакомления с образцами решений заданий студент должен выполнить свой вариант лабораторной работы. Мы не приводим ответы к задачам, дабы не лишать обучающихся удовольствия от самостоятельного решения. Заинтересовавшиеся коллеги могут получить ответы по адресу: onzhdanov@mail.ru.

КЛАССИЧЕСКИЕ ШИФРЫ

Разработкой методов преобразования *(шифрования)* информации с целью ее защиты от незаконных пользователей занимается *криптография.* Такие методы и способы преобразования информации называются *шифрами.*

Шифрование (зашифрование) — процесс применения шифра к защищаемой информации, т. е. преобразование защищаемой информации *(открытого текста)* в шифрованное сообщение *(шифртекст, криптограмму)* с помощью определенных правил, содержащихся в шифре.

Дешифрование — процесс, обратный шифрованию, т. е. преобразование шифрованного сообщения в защищаемую информацию с помощью определенных правил, содержащихся в шифре.

Криптография — прикладная наука, она использует самые последние достижения фундаментальных наук и, в первую очередь, математики. С другой стороны, все конкретные задачи криптографии существенно зависят от уровня развития техники и технологии, от применяемых средств связи и способов передачи информации.

Современная *криптография* является областью знаний, связанной с решением таких проблем безопасности информации, как конфиденциальность, целостность, аутентификация и невозможность отказа сторон от авторства. Достижение этих требований безопасности информационного взаимодействия и составляет основные цели криптографии. Они определяются следующим образом.

Обеспечение *конфиденциальности* — решение проблемы защиты информации от ознакомления с ее содержанием со стороны лиц, не имеющих права доступа к ней. В зависимости от контекста вместо термина "конфиденциальная" информация могут выступать термины "секретная", "частная", "ограниченного доступа" информация.

Обеспечение *целостности* — гарантирование невозможности несанкционированного изменения информации. Для гарантии целостности необходим простой и надежный критерий обнаружения любых манипуляций с данными. Манипуляции с данными включают вставку, удаление и замену.

Обеспечение *аутентификации* — разработка методов подтверждения подлинности сторон *(идентификация)* и самой информации в процессе информационного взаимодействия. Информация, передаваемая по каналу связи, должна быть аутентифицирована по источнику, времени создания, содержанию данных, времени пересылки и т. д.

Обеспечение *невозможности отказа от авторства* — предотвращение возможности отказа субъектов от некоторых из совершенных ими действий. Рассмотрим средства для достижения этих целей более подробно.

Традиционной задачей криптографии является проблема обеспечения конфиденциальности информации при передаче сообщений по контролируемому противником каналу связи. В простейшем случае эта задача описывается взаимодействием трех субъектов (сторон). Владелец информации, называемый обычно *отправителем,* осуществляет преобразование исходной *(открытой)*

информации (сам процесс преобразования называется *шифрованием)* в форму передаваемых *получателю*по открытому каналу связи *шифрованных* сообщений с целью ее защиты от противника.

Под *противником* понимается любой субъект, не имеющий права ознакомления с содержанием передаваемой информации. В качестве противника может выступать *криптоаналитик,* владеющий методами раскрытия шифров. Законный получатель информации осуществляет *расшифрование* полученных сообщений. Противник пытается овладеть защищаемой информацией (его действия обычно называют *атаками).* При этом он может совершать как пассивные, так и активные действия. *Пассивные* атаки связаны с прослушиванием, анализом трафика, перехватом, записью передаваемых шифрованных сообщений, *дешифрованием,* то есть попытками "взломать" защиту с целью овладения информацией.

При проведении *активных* атак противник может прерывать процесс передачи сообщений, создавать поддельные (сфабрикованные) или модифицировать передаваемые шифрованные сообщения. Эти активные действия называют попытками *имитации* и *подмены* соответственно.

Под *шифром* обычно понимается семейство обратимых преобразований, каждое из которых определяется некоторым параметром, называемым ключом, а также порядком применения данного преобразования, называемым *режимом шифрования*.

Ключ — это важнейший компонент шифра, отвечающий за выбор преобразования, применяемого для зашифрования конкретного сообщения. Обычно ключ представляет собой некоторую буквенную или числовую последовательность. Эта последовательность как бы "настраивает" алгоритм шифрования.

Каждое преобразование однозначно определяется ключом и описывается некоторым *криптографическим алгоритмом.* Один и тот же криптографический алгоритм может применяться для шифрования в различных режимах. Тем самым реализуются различные способы шифрования (простая замена, гаммирование и т. п.). Каждый режим шифрования имеет как свои преимущества, так и недостатки. Поэтому выбор режима зависит от конкретной ситуации. При расшифровании используется криптографический алгоритм, который в общем случае может отличаться от алгоритма, применяемого для зашифрования сообщения. Соответственно могут различаться ключи зашифрования и расшифрования. Пару алгоритмов зашифрования и расшифрования обычно называют *криптосистемой (шифрсистемой),* а реализующие их устройства — *шифртехникой.*

Если обозначить через *M* открытое, а через *C* шифрованное сообщения, то процессы зашифрования и расшифрования можно записать в виде равенств

$E_{k1}(M)=C$

$D_{k2}(C)=M$

в которых алгоритмы зашифрования *E* и расшифрования *D* должны удовлетворять равенству

$D_{k2}(E_{k1}(M))=M$

Наряду с конфиденциальностью не менее важной задачей является обеспечение *целостности* информации, другими словами, — неизменности ее в

процессе передачи или хранении. Решение этой задачи предполагает разработку средств, позволяющих обнаруживать не столько случайные искажения (для этой цели вполне подходят методы теории кодирования с обнаружением и исправлением ошибок), сколько целенаправленное навязывание противником ложной информации. Для этого в передаваемую информацию вносится избыточность. Как правило, это достигается добавлением к сообщению некоторой проверочной комбинации, вычисляемой с помощью специального алгоритма и играющей роль контрольной суммы для проверки целостности полученного сообщения. Главное отличие такого метода от методов теории кодирования состоит в том, что алгоритм выработки проверочной комбинации является "криптографическим", то есть зависящим от секретного ключа. Без знания секретного ключа вероятность успешного навязывания противником искаженной или ложной информации мала. Такая вероятность служит мерой *имитостойкости* шифра, то есть способности самого шифра противостоять активным атакам со стороны противника.

Итак, для проверки целостности к сообщению *M* добавляется проверочная комбинация *S,* называемая *кодом аутентификации сообщения* (сокращенно — КАС) или *имитовставкой.* В этом случае по каналу связи передается пара *C = (M, S).* При получении сообщения *M* пользователь вычисляет значение проверочной комбинации и сравнивает его с полученным контрольным значением *S* Несовпадение говорит о том, что данные были изменены.

Как правило, код аутентификации является значением некоторой (зависящей от секретного ключа) криптографической *хеш-функции* от данного сообщения: $h_k(M) = S$ К кодам аутентификации предъявляются определенные требования. К ним относятся:

— невозможность вычисления значения $h_k(M) = S$ для заданного сообщения *M* без знания ключа k,

— невозможность подбора для заданного сообщения *M* с известным значением $h_k(M)=S$ другого сообщения M_1 с известным значением h_k $(M_1) = S_1$, без знания ключа k.

Первое требование направлено против создания поддельных (сфабрикованных) сообщений при атаках типа *имитация;* второе — против модификации передаваемых сообщений при атаках типа *подмена.*

Аутентификация — установление подлинности. В общем случае этот термин может относиться ко всем аспектам информационного взаимодействия: сеансу связи, сторонам, передаваемым сообщениям и т. д.

Установление подлинности (то есть проверка и подтверждение) всех аспектов информационного взаимодействия является важной составной частью проблемы обеспечения достоверности получаемой информации. Особенно остро эта проблема стоит в случае не доверяющих друг другу сторон, когда источником угроз может служить не только третья сторона (противник), но и сторона, с которой осуществляется взаимодействие.

Применительно к сеансу связи аутентификация означает проверку: целостности соединения, невозможности повторной передачи данных противником и своевременности передачи данных. Для этого, как правило, используют

дополнительные параметры, позволяющие "сцепить" передаваемые данные в легко проверяемую последовательность. Это достигается, например, путем вставки в сообщения некоторых специальных чисел или *меток времени.* Они позволяют предотвратить попытки повторной передачи, изменения порядка следования или обратной отсылки части переданных сообщений. При этом такие вставки в передаваемом сообщении необходимо защищать (например, с помощью шифрования) от возможных подделок и искажений.

Применительно к сторонам взаимодействия аутентификация означает проверку одной из сторон того, что взаимодействующая с ней сторона — именно та, за которую она себя выдает. Часто аутентификацию сторон называют также *идентификацией*.

Основным средством для проведения идентификации являются *протоколы идентификации,* позволяющие осуществлять идентификацию (и аутентификацию) каждой из участвующих во взаимодействии и не доверяющих друг другу сторон. Различают *протоколы односторонней* и *взаимной идентификации.*

Протокол — это распределенный алгоритм, определяющий последовательность действий каждой из сторон. В процессе выполнения протокола идентификации каждая из сторон не передает никакой информации о своем секретном ключе, а хранит его у себя и использует для формирования ответных сообщений на запросы, поступающие при выполнении протокола.

Наконец, применительно к самой информации аутентификация означает проверку того, что информация, передаваемая по каналу, является подлинной по содержанию, источнику, времени создания, времени пересылки и т. д.

Проверка подлинности содержания информации сводится, по сути, к проверке ее неизменности (с момента создания) в процессе передачи или хранения, то есть проверке целостности.

Аутентификация источника данных означает подтверждение того, что исходный документ был создан именно заявленным источником.

Заметим, что если стороны доверяют друг другу и обладают общим секретным ключом, то аутентификацию сторон можно обеспечить применением кода аутентификации. Действительно, каждое успешно декодированное получателем сообщение может быть создано только отправителем, так как только он знает их общий секретный ключ. Для не доверяющих друг другу сторон решение подобных задач с использованием общего секретного ключа становится невозможным. Поэтому при аутентификации источника данных нужен механизм цифровой подписи, который будет рассмотрен ниже.

В целом, аутентификация источника данных выполняет ту же роль, что и протокол идентификации. Отличие заключается только в том, что в первом случае имеется некоторая передаваемая информация, авторство которой требуется установить, а во втором требуется просто установить сторону, с шторой осуществляется взаимодействие.

Математические модели открытого текста

Потребность в математических моделях открытого текста продиктована, прежде всего, следующими соображениями. Во-первых, даже при отсутствии ограничений на временные и материальные затраты по выявлению закономерностей, имеющих место в открытых текстах, нельзя гарантировать того, что такие свойства указаны с достаточной полнотой. Например, хорошо известно, что частотные свойства текстов в значительной степени зависят от их характера. Поэтому при математических исследованиях свойств шифров прибегают к упрощающему моделированию, в частности, реальный открытый текст заменяется его моделью, отражающей наиболее важные его свойства. Во-вторых, при автоматизации методов криптоанализа, связанных с перебором ключей, требуется "научить" ЭВМ отличать открытый текст от случайной последовательности знаков. Ясно, что соответствующий критерий может выявить лишь адекватность последовательности знаков некоторой модели открытого текста.

Один из естественных подходов к моделированию открытых текстов связан с учетом их частотных характеристик, приближения для которых можно вычислить с нужной точностью, исследуя тексты достаточной длины. Основанием для такого подхода является устойчивость частот *к* -грамм или целых словоформ реальных языков человеческого общения (то есть отдельных букв, слогов, слов и некоторых словосочетаний). Основанием для построения модели может служить также и теоретико-информационный подход, развитый в работах К. Шеннона.

Учет частот k-грамм приводит к следующей модели открытого текста. Пусть $P^{(k)}(A)$ представляет собой массив, состоящий из приближений для вероятностей $p(b_1,b_2,...,b_k)$ появления k-грамм $b_1b_2...b_k$ в открытом тексте, $k \in N$,

$A = (a_1,...,a_n)$ — алфавит открытого текста, $b_i \in A$, $i = 1,k$.

Тогда источник "открытого текста" генерирует последовательность $c_1,c_2,...,c_k,c_{k+1},...$ знаков алфавита A, в которой k-грамма $c_1c_2...c_k$ появляется с вероятностью $p(c_1c_2...c_k) \in P^{(k)}(A)$, следующая k-грамма $c_1c_2...c_{k+1}$ появляется с вероятность $p(c_2c_3...c_{k+1}) \in P^{(k)}(A)$ и т. д. Назовем построенную модель открытого текста *вероятностной моделью k-го приближения.*

Таким образом, простейшая модель открытого текста -*вероятностная модель первого приближения* – представляет собой последовательность знаков $c_1,c_2,...$, в которой каждый знак ci, $i = 1,2,...$, появляется с вероятностью $p(c_i) \in P^{(1)}(A)$, независимо от других знаков. Будем называть также эту модель *позначной моделью открытого текста.* В такой модели открытый текст $c_1c_2...c_l$ имеет вероятность

$$p(c_1c_2...c_l) = \prod_{i=1}^{l} p(c_i).$$

В вероятностной модели второго приближения первый знак c_1 имеет вероятность $p(c_1) \in P^{(1)}(A)$, а каждый следующий знак c_i зависит от предыдущего и появляется с вероятностью

$$p(c_i / c_{i-1}) = \frac{p(c_{i-1}c)}{p(c_{i-1})},$$

где $p(c_{i-1}c_i) \in P^{(2)}(A)$, $p(c_{i-1}) \in P^{(1)}(A)$, $i = 2,3,....$ Другими словами, модель открытого текста второго приближения представляет собой *простую однородную цепь Маркова.* В такой модели открытый текст $c_1c_2...c_l$ имеет вероятность

$$p(c_1c_2...c_l) = p(c_1) \cdot \prod_{i=2}^{l} p(c_i / c_{i-1}).$$

Модели открытого текста более высоких приближений учитывают зависимость каждого знака от большего числа предыдущих знаков. Ясно, что чем выше степень приближения, тем более "читаемыми" являются соответствующие модели. Проводились эксперименты по моделированию открытых текстов с помощью ЭВМ.

Отметим, что с более общих позиций открытый текст может рассматриваться как реализация *стационарного эргодического случайного процесса с дискретным временем и конечным числом состояний*.

Критерии распознавания открытого текста

Заменив реальный открытый текст его моделью, мы можем теперь построить критерий распознавания открытого текста. При этом можно воспользоваться либо стандартными методами различения статистических гипотез, либо наличием в открытых текстах некоторых запретов, таких, например, как биграмма ЪЪ в русском тексте. Проиллюстрируем первый подход при распознавании позначной модели открытого текста.

Итак, согласно нашей договоренности, открытый текст представляет собой реализацию независимых испытаний случайной величины, значениями которой являются буквы алфавита $A = \{a_1,...,a_n\}$, появляющиеся в соответствии с распределением вероятностей $P^{(1)}(A) = (p(a_1),...,p(a_n))$. Требуется ' определить, является ли случайная последовательность $c_1c_2...c_l$ букв алфавита A открытым текстом или нет.

Пусть H_0 — гипотеза, состоящая в том, что данная последовательность — открытый текст, H_1 — альтернативная гипотеза. В простейшем случае последовательность $c_1c_2...c_l$ можно рассматривать при гипотезе H_1 как случайную и равновероятную. Эта альтернатива отвечает субъективному представлению о том, что при расшифровании криптограммы с помощью ложного ключа получается "бессмысленная" последовательность знаков. В более общем случае можно считать, что при гипотезе H1 последовательность $c_1c_2...c_l$ представляет собой реализацию независимых испытаний некоторой случайной величины, значениями которой являются буквы алфавита $A = \{a_1,...,a_n\}$, появляющиеся в соответствии с распределением вероятностей $Q^{(1)}(A) = (q(a_l),...,q(a_n))$. При таких договоренностях можно применить, например, *наиболее мощный критерий* различения двух простых гипотез, который дает *лемма Неймана—Пирсона*.

В силу своего вероятностного характера такой критерий может совершать ошибки двух родов. Критерий может принять открытый текст за случайный набор знаков. Такая ошибка обычно называется *ошибкой первого рода,* ее вероятность равна $\alpha = p\{H_1/H_0\}$. Аналогично вводится *ошибка второго рода* и ее вероятность $\beta = p\{H_0/H_1\}$. Эти ошибки определяют качество работы критерия. В криптографических исследованиях естественно минимизировать вероятность

ошибки первого рода, чтобы не "пропустить" открытый текст. Лемма Неймана—Пирсона при заданной вероятности первого рода минимизирует также вероятность ошибки второго рода.

Критерии на открытый текст, использующие запретные сочетания знаков, например *к* -граммы подряд идущих букв, будем называть *критериями запретных k-грамм.* Они устроены чрезвычайно просто. Отбирается некоторое число *s* редких *k*-грамм, которые объявляются запретными. Теперь, просматривая последовательно *k*-грамму за *k*-граммой анализируемой последовательности $c_1c_2...c_l$, мы объявляем ее случайной, как только в ней встретится одна из запретных *k*-грамм, и открытым текстом в противном случае. Такие критерии также могут совершать ошибки в принятии решения. В простейших случаях их можно рассчитать. Несмотря на свою простоту, критерии запретных *k*-грамм являются весьма эффективными.

Классификация шифров

В качестве первичного признака, по которому производится классификация шифров, используется тип преобразования, осуществляемого с открытым текстом при шифровании. Если фрагменты открытого текста (отдельные буквы или группы букв) заменяются некоторыми их эквивалентами в шифртексте, то соответствующий шифр относится к классу *шифров замены.* Если буквы открытого текста при шифровании лишь меняются местами друг с другом, то мы имеем дело с *шифром перестановки.* С целью повышения надежности шифрования шифрованный текст, полученный применением некоторого шифра, может быть еще раз зашифрован с помощью другого шифра. Всевозможные такие композиции различных шифров приводят к третьему классу шифров, которые обычно называют *композиционными шифрами.* Заметим, что композиционный шифр может не входить ни в класс шифров замены, ни в класс шифров перестановки (рис. 1).

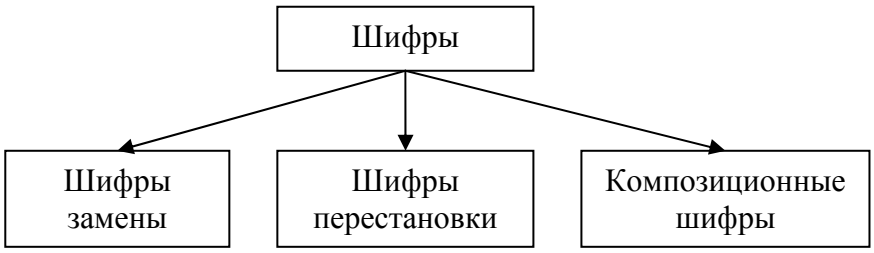


Рисунок 1. Классификация шифров

Шифры перестановки

Шифры перестановки, или транспозиции, изменяют только порядок следования символов или других элементов исходного текста. Классическим примером такого шифра является система, использующая карточку с отверстиями – *решетку Кардано,* которая при наложении на лист бумаги оставляет открытыми лишь некоторые его части. При зашифровке буквы сообщения вписываются в эти отверстия. При расшифровке сообщение вписывается в диаграмму нужных размеров, затем накладывается решетка, после чего на виду оказываются только буквы открытого текста.

Решетки можно использовать двумя различными способами. В первом случае

зашифрованный текст состоит только из букв исходного сообщения. Решетка изготавливается таким образом, чтобы при ее последовательном использовании в различных положениях каждая клетка лежащего под ней листа бумаги оказалась занятой. Примером такой решетки является *поворотная решетка*, показанная на рис.1. Если такую решетку последовательно поворачивать на 90° после заполнения всех открытых при данном положении клеток, то при возврате решетки в исходное положение все клетки окажутся заполненными. Числа, стоящие в клетках, облегчают изготовление решетки. В каждом из концентрических окаймлений должна быть вырезана только одна клетка из тех, которые имеют одинаковый номер. Второй, стеганографический метод использования решетки позволяет скрыть факт передачи секретного сообщения. В этом случае заполняется только часть листа бумаги, лежащего под решеткой, после чего буквы или слова исходного текста окружаются ложным текстом.

1	2	3	4	5	1
5	1	2	3	1	2
4	3	1	1	2	3
3	2	1	1	3	4
2	1	3	2	1	5
1	5	4	3	2	1

Рисунок 2. Пример поворотной решетки

Рассмотрим усложненную перестановку по таблице. Пример таблицы для реализации этого метода шифрования показан на рис.3. Таблица представляет собой матрицу размерностью 6 х 6, в которую построчно вписывается искомое сообщение. При считывании информации по столбцам в соответствии с последовательностью чисел ключа получается шифротекст. Усложнение заключается в том, что некоторые ячейки таблицы не используются. При зашифровании сообщения

КОМАНДОВАТЬ ПАРАДОМ БУДУ Я

получим:

ОЬБНАОДКДМУМВ АУ ОТР ААПДЯ,

Ключ					
2	4	0	3	5	1
К	О		М	А	Н
Д		О	В	А	
	Т	Ь		П	А

	Р		А	Д	О
М		Б	У		Д
У				Я	

Рисунок 3. Пример шифрования методом усложненной перестановки по таблице

При расшифровании буквы шифротекста записываются по столбцам в соответствии с последовательностью чисел ключа, после чего исходный текст считывается по строкам. Для удобства запоминания ключа применяют перестановку столбцов таблицы по ключевому слову или фразе, всем символам которых ставятся в соответствие номера, определяемые порядком соответствующих букв в алфавите. Например, при выборе в качестве ключа слова ИНГОДА последовательность использования столбцов будет иметь вид 462531.

Также возможны и другие варианты шифра перестановки, например, шифры столбцовой и двойной перестановки.

Шифры замены

Большое влияние на развитие криптографии оказали появившиеся в середине XX века работы американского математика Клода Шеннона. В этих работах были заложены основы теории информации, а также был разработан математический аппарат для исследований во многих областях науки, связанных с информацией. Более того, принято считать, что теория информации как наука родилась в 1948 году после публикации работы К. Шеннона «Математическая теория связи» (см. приложение).

В своей работе «Теория связи в секретных системах» Клод Шеннон обобщил накопленный до него опыт разработки шифров. Оказалось, что даже в очень сложных шифрах в качестве типичных компонентов можно выделить такие простые шифры как *шифры замены, шифры перестановки* или их сочетания.

Шифр замены является простейшим, наиболее популярным шифром. Типичными примерами являются шифр Цезаря, «цифирная азбука» Петра Великого и «пляшущие человечки» А. Конан Дойла. Как видно из самого названия, шифр замены осуществляет преобразование замены букв или других «частей» открытого текста на аналогичные «части» шифрованного текста. Легко дать математическое описание шифра замены. Пусть X *и* Y – два алфавита (открытого и шифрованного текстов соответственно), состоящие из одинакового числа символов. Пусть также $g: X \longrightarrow Y$ — взаимнооднозначное отображение X в Y. Тогда шифр замены действует так: открытый текст $x_1x_2...x_n$ преобразуется в шифрованный текст $g(x_1)g(x_2)... g(x_n)$.

Шифр перестановки, как видно из названия, осуществляет преобразование перестановки букв в открытом тексте. Типичным примером шифра перестановки является шифр «Сцитала». Обычно открытый текст разбивается на отрезки равной длины и каждый отрезок шифруется независимо. Пусть, например, длина отрезков

равна n и σ — взаимнооднозначное отображение множества *{1,2,..., n}* в себя. Тогда шифр перестановки действует так: отрезок открытого текста $x_1...x_n$ преобразуется в отрезок шифрованного текста

Математическая модель шифра замены

Определим модель Σ_A *=(X,K,Y,E,D)* произвольного шифра замены. Будем считать, что открытые и шифрованные тексты являются словами в алфавитах *A и B* соответственно: $X \subset A^*$, $Y \subset B^*$, $|A|=n$, $|B| = m$. Здесь и далее C* обозначает множество слов конечной длины в алфавите C.

Перед зашифрованием открытый текст предварительно представляется в виде последовательности подслов, называемых *шифрвеличинами.* При зашифровании шифрвеличины заменяются некоторыми их эквивалентами в шифртексте, которые назовем *шифробозначениями.* Как шифрвеличины, так и шифробозначения представляют собой слова из A^* и B^* соответственно.

Пусть $U = \{u_1,..,u_N\}$ — множество возможных шифрвеличин, $V = \{v_1,...,v_M\}$ — множество возможных шифробозначений. Эти множества должны быть такими, чтобы любые тексты $x \in X$, $y \in Y$ можно было представить словами из U^*, V^* соответственно. Требование однозначности расшифрования влечет неравенства $N \geq n$, $M \geq m$, $M \geq N$. Для определения правила зашифрования E_k(x) в общем случае нам понадобится ряд обозначений и понятие *распределителя,* который, по сути, и будет выбирать в каждом такте шифрования замену соответствующей шифрвеличине.

Поскольку $M \geq N$, множество V можно представить в виде объединения $V = \bigcup_{i=1}^{N} V_\alpha^{(i)}$ непересекающихся непустых подмножеств $V^{(i)}$. Рассмотрим произвольное семейство, состоящее из r таких разбиений множества V:

$$V = \bigcup_{i=1}^{N} V_\alpha^{(i)},\ \alpha = \overline{1,r},\ r \in N,$$

и соответствующее семейство биекций

$$\varphi_\alpha : U \to \{V_\alpha^{(1)},...,V_\alpha^{N}\},$$

для которых $\varphi_\alpha(u_i) = V_\alpha^{(i)},\ i = \overline{1,N}$.

Рассмотрим также произвольное отображение $\psi : K \times N \to N_r^*$, где $N_r = \{1,2,...,r\}$, такое, что для любых $k \in K, l \in N$

$$\psi(k,l) = \alpha_1^{(k)}...\alpha_l^{(k)},\ \alpha_j^{(k)} \in N_r,\quad j = \overline{1,l}.$$

Назовем последовательность $\psi(к,l)$ *распределителем,* отвечающим данным значениям $к \in K$, $l \in$ N.

Теперь мы сможем определить правило зашифрования произвольного шифра замены. Пусть

$$x \in X,\ x = x_1...x_l,\ x_i \in U,\ i = 1,l;\ k \in K$$

и $\psi(к,I) = a_1^{(k)}...a_1^{(k)}$. Тогда $E_к(x) = y$, где $y = y_1...y_l$,

$y_j = \varphi_{\alpha_j^{(k)}}(x),\ j = \overline{1,l}.$

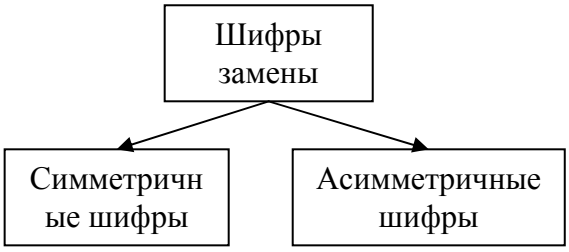
В качестве y_j можно выбрать любой элемент множества m $\varphi_{\alpha_j^{(k)}}(x_j)$. Всякий раз при шифровании этот выбор можно производить случайно, например, с помощью некоторого *рандомизатора* типа игровой рулетки. Подчеркнем, что такая

многозначность при зашифровании не препятствует расшифрованию, так как $V_\alpha^{(i)} \bigcap V_\alpha^{(j)} = \emptyset$ при $i \neq j$.

Классификация шифров замены

Если ключ зашифрования совпадает с ключом расшифрования: $k_3 = k_p$, то такие шифры называют *симметричными,* если же $k_3 \neq k_p$ — *асимметричными.*

В связи с указанным различием в использовании ключей сделаем еще один шаг в классификации:



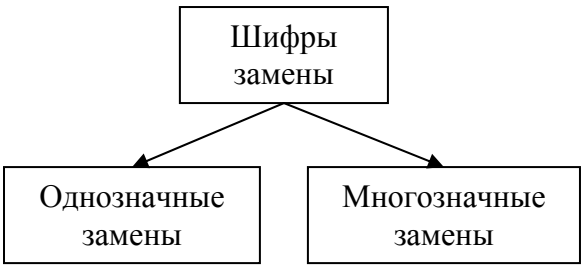
Отметим также, что в приведенном определении правило зашифрования $E_k(x)$ является, вообще говоря, *многозначной функцией.* Выбор ее значений представляет собой некоторую проблему, которая делает многозначные функции $E_k(x)$ не слишком удобными для использования. Избавиться от этой проблемы позволяет использование однозначных функций, что приводит к естественному разделению всех шифров замены на *однозначные* и *многозначные замены* (называемых также в литературе *омофонами).*

Для однозначных шифров замены справедливо свойство:

$$\forall \alpha, i : \left| V_\alpha^{(i)} \right| = 1;$$

для многозначных шифров замены:

$$\exists \alpha, i : \left| V_\alpha^{(i)} \right| > 1;$$



Исторически известный шифр — *пропорциональной замены* представляет собой пример шифра многозначной замены, *шифр гаммирования* - пример шифра однозначной замены. Далее мы будем заниматься в основном изучением однозначных замен, получивших наибольшее практическое применение. Итак, далее $M = N$ и $\varphi_\alpha(u_i) = v_{\alpha,i}$, $i = \overline{1, M}$.

Заметим, что правило зашифрования E_k естественным образом индуцирует отображение $\widetilde{E}_k : U \to V$, которое в свою очередь продолжается до отображения $\widetilde{E}_k : U^* \to V^*$. Для упрощения записи будем использовать одно обозначение E_k для каждого из трех указанных отображений.

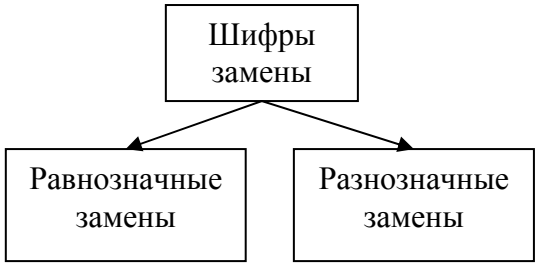
В силу инъективности (по k) отображения E_k и того, что $|U| = |V|$, введенные в общем случае отображения φ_α являются биекциями $\varphi_\alpha : U \leftrightarrow V$, определенными равенствами $\varphi_\alpha(u_i) = v_\alpha^{(i)}, i = \overline{1,N}, \alpha = \overline{1,r}.$. Число таких биекций не превосходит $N!$.

Для шифра однозначной замены определение правила зашифрования можно уточнить: в формуле включение следует заменить равенством

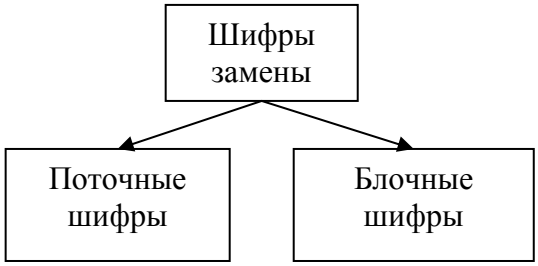
$$y_j = \varphi_{\alpha_j^{(k)}}(x_j), \quad j = \overline{1,l}.$$

Введем еще ряд определений.

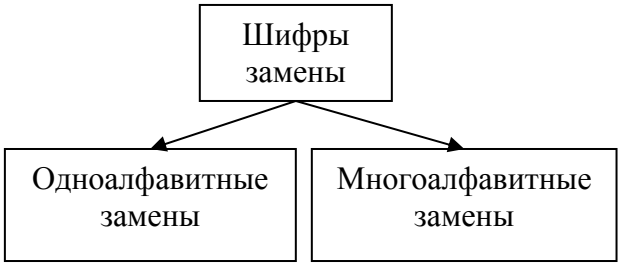
Если для некоторого числа $q \in \mathrm{N}$ выполняются включения $v_i \in B^q$, $i=1,N$, то соответствующий шифр замены будем называть *шифром равнозначной замены.* В противном случае — *шифром разнозначной замены:*



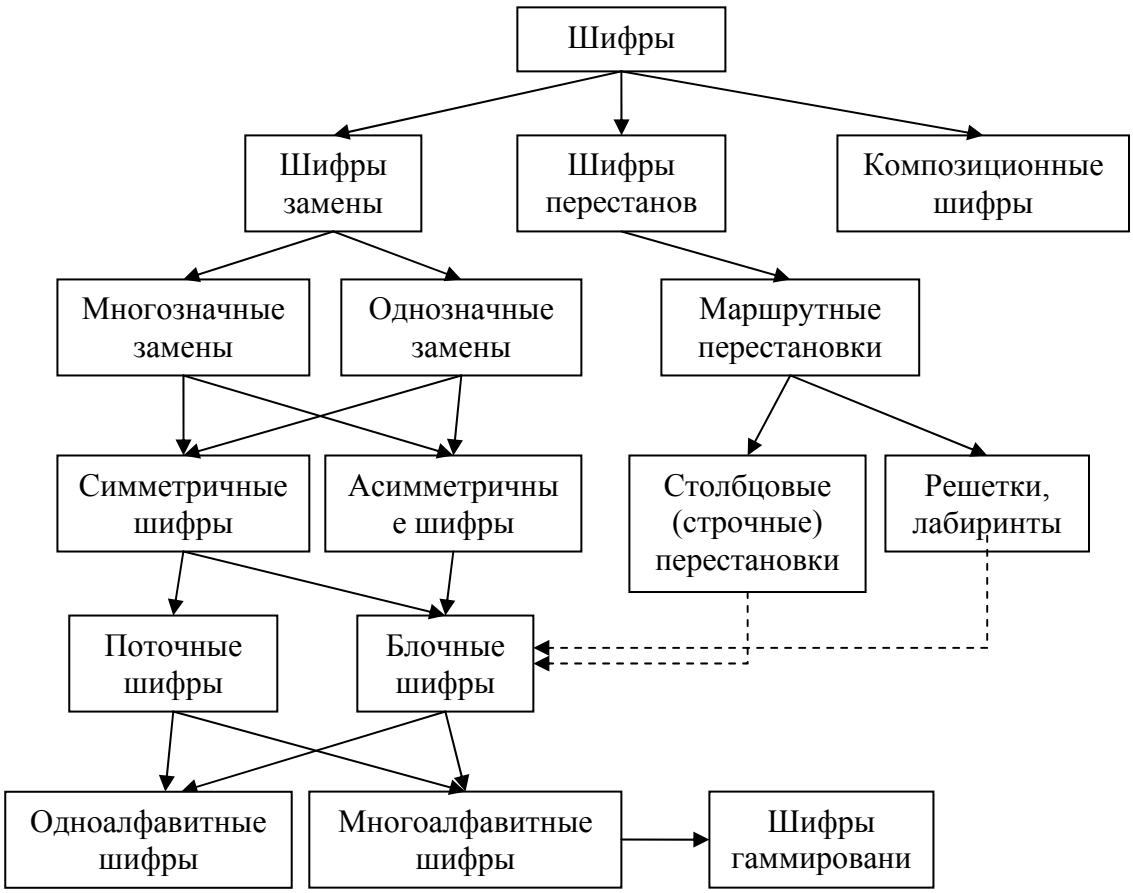
В подавляющем большинстве случаев используются шифры замены, для которых $U \in A^p$, для некоторого $p \in \mathrm{N}$. При $p = 1$ говорят о *поточных шифрах замены,* при $p > 1$ — о *блочных шифрах замены:*



Следующее определение. В случае r = 1 шифр замены называют *одноалфавитным шифром замены* или *шифром простой замены.* В противном случае – *многоалфавитным шифром замены:*



Ограничиваясь наиболее важными классами шифров замены и исторически известными классами шифров перестановки, сведем результаты классификации в схему, изображенную на рисунке.



Следует подчеркнуть, что стрелки, выходящие из любого прямоугольника схемы, указывают лишь на наиболее значимые частные подклассы шифров. Пунктирные стрелки, ведущие из подклассов шифров перестановки, означают, что эти шифры можно рассматривать и как блочные шифры замены в соответствии с тем, что открытый текст делится при шифровании на блоки фиксированной длины, в каждом из которых производится некоторая перестановка букв. Одноалфавитные и многоалфавитные шифры могут быть как поточными, так и блочными. В то же время шифры гаммирования, образующие подкласс многоалфавитных шифров, относятся к поточным, а не к блочным шифрам. Кроме того, они являются симметричными, а не асимметричными шифрами.

Шифр Виженера

Наиболее известными являются шифры замены, или подстановки, особенностью которых является замена символов (или слов, или других частей сообщения) открытого текста соответствующими символами, принадлежащими алфавиту шифротекста. Различают *одноалфавитную* и *многоалфавитную* замену. Вскрытие одноалфавитных шифров основано на учете частоты появления отдельных букв или их сочетаний (биграмм, триграмм и т. п.) в данном языке. Классические примеры вскрытия таких шифров содержатся в рассказах Э. По "Золотой жук" и А.Конан Дойля "Пляшущие человечки".

Примером многоалфавитного шифра замены является так называемая система Виженера. Шифрование осуществляется по таблице, представляющей собой квадратную матрицу размерностью п X n, где п - число символов используемого

алфавита. На рис.4 показана таблица Виженера для русского языка (алфавит Z_{32}- 32 буквы и пробел). Первая строка содержит все символы алфавита. Каждая следующая строка получается из предыдущей циклическим сдвигом последней на символ влево.

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

Рисунок 4. Таблица Виженера для алфавита Z_{32}

Выбирается ключ или ключевая фраза. После чего процесс зашифрования осуществляется следующим образом. Под каждой буквой исходного сообщения последовательно записываются буквы ключа; если ключ оказался короче сообщения, его используют несколько раз. Каждая буква шифротекста находится на пересечении столбца таблицы, определяемого буквой открытого текста, и строки, определяемой буквой ключа. Пусть, например, требуется зашифровать сообщение:

ГРУЗИТЕ АПЕЛЬСИНЫ БОЧКАМИ ТЧК БРАТЬЯ КАРАМАЗОВЫ ТЧК

С помощью ключа ВЕНТИЛЬ запишем строку исходного текста с расположенной под ней строкой с циклически повторяемым ключом:

```
ГРУЗИТЕ АПЕЛЬСИНЫ БОЧКАМИ ТЧК БРАТЬЯ КАРАМАЗОВЫ ТЧК
ВЕНТИЛЬВЕНТИЛЬВЕНТИЛЬВЕНТИЛЬВЕНТИЛЬВЕНТИЛЬВЕНТИЛЬВЕ
```

В результате зашифрования, начальный этап которого показан на рисунке 5, получим шифротекст:

```
ЕХ ЩРЭАБЕЫЧУДККТИСЙЩРМЕЩЬЗЭРМДОБИЭУАДЧТШЛЕВМЪФГКЛЩП
```

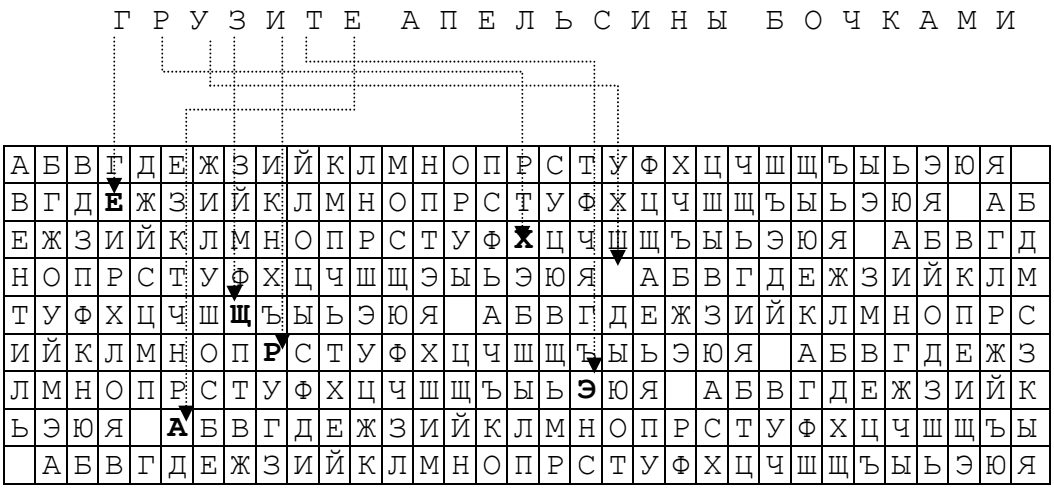


Рисунок 5. Принцип шифрования по таблице Виженера

Расшифрование осуществляется следующим образом. Под буквами шифротекста последовательно записываются буквы ключа; в строке таблицы, соответствующей очередной букве ключа, происходит поиск соответствующей буквы шифротекста. Находящаяся над ней в первой строке таблицы буква является соответствующей буквой исходного текста.

Для увеличения надежности шифра можно рекомендовать его использование после предварительной псевдослучайной перестановки букв в каждой строке таблицы. Возможны и другие модификации метода.

СОВЕТЫ ПО ВЫПОЛНЕНИЮ ЧАСТОТНОГО АНАЛИЗА АНГЛИЙСКИХ ТЕКСТОВ

(1) Начните с подсчета частоты появления каждой из букв шифр-текста. Примерно пять букв должны появляться с частотой менее 1 процента, и они вероятно, представляют собой j, k, q, x и z. Одна из букв должна появляться с частотой более 10 процентов, и она, по-видимому, представляет собой e. Если шифр-текст не подчиняется этому распределению частот, то, возможно, исходное сообщение написано не на английском языке. Вы можете определить, какой это язык, если проанализируете частотное распределение букв в шифр-тексте. К

примеру, в итальянском языке обычно есть, три буквы с частотностью более 10 процентов и 9 букв с частотностью менее 1 процента. В немецком языке буква e имеет чрезвычайно высокую частотность – 19 процентов, поэтому любой шифр-текст, в котором одни из букв встречается столь же часто, является, вполне возможно, немецким. После того как вы определили язык, для выполнения частотного анализа вам следует воспользоваться соответствующей таблицей частотности букв для данного языка. Если у вас есть нужная таблица частотности букв, то нередко удается дешифровать даже шифр-тексты на неизвестном языке.

(2) Если установлена взаимосвязь с английским языком, но, как часто и происходит, сразу же открытый текст не появляется, тогда обратите внимание на пары повторяющихся букв. В английском языке чаще всего повторяющимися буквами будут **ss, ee, tt, ff, 11, mm** и **oo**. Если в шифр-тексте имеются какие-либо повторяющиеся символы, то вы можете считать, что они представляют собой одну из этих пар.

(3) Если в шифр-тексте имеются пробелы между словами, то постарайтесь определить слова, состоящие из одной, двух или трех букв. Единственными словами в английском языке, состоящими из одной буквы, являются **a** и **I.** Чаще всего встречающимися двухбуквенными словами будут **of, to, in, it, is, be, as, at, so, we, he, by, or, on, do, if, me, my, up, an, go, no, us, am.** Наиболее часто появляющиеся трехбуквенные слова – **the** и **and**.

(4) Если удастся, подготовьте таблицу частотности букв для сообщения, которое вы стараетесь дешифровать. Например, в военных донесениях стремятся опускать местоимения и артикли, и отсутствие таких слов, как **I**, **he**, **a** и **the**, будет снижать частотность некоторых из чаще всего встречающихся букв. Если вы знаете, что работаете с военным донесением, вам следует использовать таблицу частотности букв, созданную на основе других военных донесений.

(5) Одно из самых полезных для криптоаналитика умений – это способность благодаря собственному опыту или чисто интуитивно – распознавать слова или даже целые фразы. Аль-Халил, один из первых арабских криптоаналитиков, продемонстрировал свои способности, когда взломал греческий шифр-текст. Он предположил, что шифр-текст начинается с приветствия «Во имя бога». Установив, что эти буквы соответствуют определенному фрагменту шифр-текста, он смог использовать их в качестве лома и раскрыть остальной шифр-текст. Это получило название криб.

(6) В некоторых случаях наиболее часто встречающейся буквой в шифр-тексте может быть **E**, следующей по частоте появления – **T** и так далее. Другими словами, частотность букв в шифр-тексте уже совпадает с частотностью букв в таблице. По-видимому, буква **E** в шифр-тексте является действительно **e**, и то же самое, похоже, справедливо и для других букв, и все же шифр-текст выглядит тарабарщиной. В этом случае вы столкнулись не с шифром замены, а с шифром перестановки. Все буквы остались теми же самыми, но находятся они не на своих местах.

ЗАДАНИЯ НА КРИПТОАНАЛИЗ КЛАССИЧЕСКИХ ШИФРОВ

1. ШИФР СТОЛБЦОВОЙ ПЕРЕСТАНОВКИ

При решении заданий на криптоанализ шифров перестановки необходимо восстановить начальный порядок следования букв текста. Для этого используется анализ совместимости символов, в чем может помочь таблица сочетаемости.

Таблица 1. Сочетаемость букв русского языка

Г	С	Слева		Справа	Г	С
3	97	л, д, к, т, в, р, н	**А**	л, н, с, т, р, в, к, м	12	88
80	20	я, е, у, и, а, о	**Б**	о, ы, е, а, р, у	81	19
68	32	я, т, а, е, и, о	**В**	о, а, и, ы, с, н, л, р	60	40
78	22	р, у, а, и, е, о	**Г**	о, а, р, л, и, в	69	31
72	28	р, я, у, а, и, е, о	**Д**	е, а, и, о, н, у, р, в	68	32
19	81	м, и, л, д, т, р, н	**Е**	н, т, р, с, л, в, м, и	12	88
83	17	р, е, и, а, у, о	**Ж**	е, и, д, а, н	71	29
89	11	о, е, а, и	**З**	а, н, в, о, м, д	51	49
27	73	р, т, м, и, о, л, н	**И**	с, н, в, и, е, м, к, з	25	75
55	45	ь, в, е, о, а, и, с	**К**	о, а, и, р, у, т, л, е	73	27
77	23	г, в, ы, и, е, о, а	**Л**	и, е, о, а, ь, я, ю, у	75	25
80	20	я, ы, а, и, е, о	**М**	и, е, о, у, а, н, п, ы	73	27
55	45	д, ь, н, о, а, и, е	**Н**	о, а, и, е, ы, н, у	80	20
11	89	р, п, к, в, т, н	**О**	в, с, т, р, и, д, н, м	15	85
65	35	в, с, у, а, и, е, о	**П**	о, р, е, а, у, и, л	68	32
55	45	и, к, т, а, п, о, е	**Р**	а, е, о, и, у, я ,ы, н	80	20
69	31	с, т, в, а, е, и, о	**С**	т, к, о, я, е, ь, с, н	32	68
57	43	ч, у, и, а, е, о, с	**Т**	о, а, е, и, ь, в, р, с	63	37
15	85	п, т, к, д, н, м, р	**У**	т, п, с, д, н, ю, ж	16	84
70	30	н, а, е, о, и	**Ф**	и, е, о, а, е, о, а	81	19
90	10	у, е, о, а, ы, и	**Х**	о, и, с, н, в, п, р	43	57
69	31	е, ю, н, а, и	**Ц**	и, е, а, ы	93	7
82	18	е, а, у, и, о	**Ч**	е, и, т, н	66	34
67	33	ь, у, ы, е, о, а, и, в	**Ш**	е, и, н, а, о, л	68	32
84	16	е, б, а, я, ю	**Щ**	е, и, а	97	3
0	100	м, р, т, с, б, в, н	**Ы**	л, х, е, м, и, в, с, н	56	44
0	100	н, с, т, л	**Ь**	н, к, в, п, с, е, о, и	24	76
14	86	с, ы, м, л, д, т, р, н	**Э**	н, т, р, с, к	0	100
58	42	ь, о, а, и, л, у	**Ю**	д, т, щ, ц, н, п	11	89
43	57	о, н, р, л, а, и, с	**Я**	в, с, т, п, д, к, м, л	16	84

Таблица 2. Сочетаемость букв английского языка

Г	С	Слева		Справа	Г	С
19	81	l,c,d,m,n,s,w,t,r,e,h	**A**	n,t,s,r,l,d,c,m	6	94
55	45	y,b,n,t,u,d,o,s,a,e	**B**	e,l,u,o,a,y,b,r	70	30

61	39	u,o,s,n,a,i,l,e	**C**	h,o,e,a,i,t,r,l,k	59	41
52	48	r,i,l,a,n,e	**D**	e,i,t,a,o,u	54	46
8	92	c,b,e,m,v,d,s,l,n,t,r,h	**E**	r,d,s,n,a,t,m,e,c,o	21	79
69	31	s,n,f,d,a,i,e,o	**F**	t,o,e,i,a,r,f,u	52	48
36	64	o,d,u,r,i,e,a,n	**G**	e.h.o.r.a.t.f.w.i.s	42	58
7	93	g,e,w,s,c,t	**H**	e,a,i,o	90	10
13	87	f,m,w,e,n,l,d,s,r,h,t	**I**	n,t,s,o,c,r,e,m,a,l	17	83
28	72	y,w,t,s,n,e,c,b,a,c	**J**	u,o,a,e,m,w	88	12
53	47	y,u,i,n,a,r,o,c	**К**	e,i,n,a,t,s	68	32
52	48	m,p,t,i,b,u,o,e,l,a	**L**	e,i,y,o,a,d,u	65	35
69	31	s,d,m,r,i,a,o,e	**M**	e,a,o,i,p,m	71	29
89	11	u,e,o,a,i	**N**	d,t,g,e,a,s,o,i,c	32	68
21	79	o,d,l,p,h,n,e,c,f,s,i,r,t	**O**	n,f,r,u,t,m,l,s,w,o	18	82
47	53	r,l,t,n,i,p,m,a,o,u,e,s	**P**	o,e,a,r,l,u,p,t,i,s	59	41
20	80	o,n,l,e,d,r,s	**Q**	u	100	0
70	30	p,i,u,t,a,o,e	**R**	e,o,a,t,i,s,y	61	39
48	52	d,t,o,u,r,n,s,i,a,e	**S**	t,e,o,i,s,a,h,p,u	41	59
43	57	u,o,d,t,f,e,i,n,s.a	**T**	h,i,o,e,a,t,r	38	62
35	65	p,f,t,l,b,d,s,o	**И**	n,s,t,r,l,p,b,c	8	92
88	12	r,u,o,a,i,e	**V**	e,i,o,a	99	1
48	52	g,d,y,n,s,t,o,e	**W**	a,h,i,e,o,n	80	20
95	5	u,n,i,e	**X**	p,t,i,a,u,c,k,o	38	62
24	76	b,n,a,t,e,r,l	**Y**	a,o,s,t,w,h,i,e,d,m	38	62
88	12	o,n,a,i	**Z**	e,i,w	86	14

При анализе сочетаемости букв друг с другом следует иметь в виду зависимость появления букв в открытом тексте от значительного числа предшествующих букв. Для анализа этих закономерностей используют понятие условной вероятности.

Систематически вопрос о зависимости букв алфавита в открытом тексте от предыдущих букв исследовался известным русским математиком А.А.Марковым (1856 — 1922). Он доказал, что появления букв в открытом тексте нельзя считать независимыми друг от друга. В связи с этим А. А. Марковым отмечена еще одна устойчивая закономерность открытых текстов, связанная с чередованием гласных и согласных букв. Им были подсчитаны частоты встречаемости биграмм вида гласная-гласная (*г,г*), гласная-согласная (*г,с*), согласная-гласная (*с,г*), согласная-согласная (*с,с*) в русском тексте длиной в 10^5 знаков. Результаты подсчета отражены в следующей таблице:

Таблица 3. Чередование гласных и согласных

	Г	С	Всего
Г	6588	38310	44898
С	38296	16806	55102

Пример решения:

Дан шифр-текст: СВПООЗЛУЙЬСТЬ_ЕДПСОКОКАЙЗО

Текст содержит 25 символов, что позволяет записать его в квадратную матрицу 5х5. Известно, что шифрование производилось по столбцам, следовательно, расшифрование следует проводить, меняя порядок столбцов.

С	В	П	О	О
З	Л	У	Й	Ь
С	Т	Ь	_	Е
Д	П	С	О	К
К	А	Й	З	О

Необходимо произвести анализ совместимости символов (Таблица сочетаемости букв русского и английского алфавита, а также таблицы частот биграмм представлена выше). В первом и третьем столбце сочетание СП является крайне маловероятным для русского языка, следовательно, такая последовательность столбцов быть не может. Рассмотрим другие запрещенные и маловероятные сочетания букв: ВП (2,3 столбцы), ПС (3,1 столбцы), ПВ (3,2 столбцы). Перебрав их все, получаем наиболее вероятные сочетания биграмм по столбцам:

В	О	С	П	О
Л	Ь	З	У	Й
Т	Е	С	Ь	_
П	О	Д	С	К
А	З	К	О	Й

Получаем осмысленный текст: ВОСПОЛЬЗУЙТЕСЬ_ПОДСКАЗКОЙ

Задание: Расшифровать фразу, зашифрованную столбцовой перестановкой.

1. ОКЕСНВРП_ЫРЕАДЕЫН_В_РСИКО
2. ДСЛИЕЗТЕА_Ь_ЛЬЮВМИ_ _АОЧХК
3. НМВИАИ_НЕВЕ_СМСТУОРДИАНКМ
4. ЕДСЗЬНДЕ_МУБД_УЭ_КРЗЕМНАЫ
5. СОНРЧОУО_ХДТ_ИЕИ_ВЗКАТРРИ
6. _ОНКА_БНЫЕЦВЛЕ_К_ТГОАНЕИР
7. НЗМАЕЕАА_Г_НОТВОССОТЬЯАЛС
8. РППОЕААДТВЛ_ЕБЬЛНЫЕ_ПА_ВР
9. ОПЗДЕП_ИХРДОТ_И_ВРИТЧ_САА
10. ВКЫОСИРЙУ_ОЬВНЕ_СОАПНИОТС
11. ПКТИРАОЛНАОИЧ_З_ЕСЬНЕЛНЖО

12.ИПКСОЕ_ТСМНАЧИ_ОЕН_ГДЕЛА_
13.АМВИННЬТЛЕАНЕ_ЙОВ_ОПХАРТО
14.АРЫКЗЫ_КЙТНЛ_ААЫ_ОЛБКЫТРТ
15._ПАРИИВИАРЗ_БРА_ИСТЬЛТОЕК
16.П_ЛНАЭУВКАА_ЦИИВР_ОКЧЕДРО
17.ЖВНОАН_АТЗОЬСН_ЫО_ФВИИКИЗ
18.ОТВГОСЕЬЬТАДВ_С_ЬЗАТТЕЫАЧ
19.ЯАМРИТ_ДЖЕХ_СВЕД_ТСУВЕТНО
20.УЬБДТ_ОЕГТВ_ОЫКЭА_ВКАИУЦИ
21.ЛТБЕЧЛЖЫЕ_ _ОАПТЖРДУ_ЛМНОА
22.ИТПРКРФАГО_АВЯИА_ЯНЖУАКАН
23.ПКЕЕРРПО_ЙУСТ_ИТПСУТЛЯЕИН
24.ИЬЖЗНСД_ТДН_ЕТ_НУВЕУРЫГОЫ
25.ЕОУРВА_НЬРИАДИЦЕПИ_РНШВЫЕ

2. ШИФР ДВОЙНОЙ ПЕРЕСТАНОВКИ

Пример решения:

Дан шифр-текст: ЫОЕЧТТОУ_СНСОРЧТРНАИДЬН_Е

Текст содержит 25 символов, что позволяет записать его в квадратную матрицу 5х5. известно, что шифрование производилось сначала по столбцам, а затем по строкам, следовательно, расшифрование следует проводить тем же способом.

Ы	О	Е	Ч	Т
Т	О	У	_	С
Н	С	О	Р	Ч
Т	Р	Н	А	И
Д	Ь	Н	_	Е

Производим анализ совместимости символов. Если в примере столюцовой перестановки можно было легко подобрать нужную комбинацию путем перебора, то здесь лучше воспользоваться таблицей частот букв русского языка (см. приложение). Для оптимизации скорости выполнения задания можно проверить все комбинации букв только в первой строке. Получаем ОЕ-15, ОЧ-12, ЕТ-33, ТЕ-31, ЧО-х, ЕО-7, ЧЫ-х, ОЫ-х, ТЫ-11, ТЧ-1, ЧЕ-23 (где х-запрещенная комбинация).

Из полученных результатов можно предположить следующую комбинацию замены столбцов **2 4 3 5 1:**

О	Ч	Е	Т	Ы
О	_	У	С	Т
С	Р	О	Ч	Н

Р	А	Н	И	Т
Ь	_	Н	Е	Д

Теперь необходимо переставить строки в нужном порядке. **3 2 4 5 1:**

С	Р	О	Ч	Н
О	_	У	С	Т
Р	А	Н	И	Т
Ь	_	Н	Е	Д
О	Ч	Е	Т	Ы

Получаем осмысленный текст: СРОЧНО_УСТРАНИТЬ_НЕДОЧЕТЫ

Задание: Расшифровать фразу, зашифрованную двойной перестановкой (сначала были переставлены столбцы, затем строки)

1. СЯСЕ_ _ЛУНЫИАККННОГЯДУЧАТН
2. МСЕЫ_ЛЫВЕНТОСАНТУЕИ_РЛПОБ
3. АМНРИД_УЕБСЫ_ЕЙРСООКОТНВ_
4. ОПЧУЛС_БООНЕВ_ОЖАЕОНЕЩЕИН
5. ЕШИАНИРЛПГЕЧАВРВ_СЕЫНА_ЛО
6. АРАВНРСВЕЕОАВ_ЗАНЯА_КМРЕИ
7. А_ЛТАВЙООЛСО_ТВ_ШЕЕНЕСТ_Ь
8. ФИ_ЗИММУЫНУУБК_Е_ДЬШЫИВЧУ
9. ВР_ЕСДЕИ_ТПХРОИ_ЗБУАДНУА_
10. ЦТААЙПЕЕ_ТБГУРРСВЬЕ_ОРЗВВ
11. АВАРНСЧАА_НЕДВЕДЕРПЕОЙ_ИС
12. ДОПК_СОПАЛЕЧНЛ_ГИНЙОИЖЕ_Т
13. ЛУАЗИЯНСА_ДТДЕАИ_ШРФЕОНГ_
14. С_ОЯНВ_СЬСЛААВРЧЕАРТОГДЕС
15. ЗШАФИПРАЛОЕНЖ_ОЬН_ДАРВОНА
16. КЭЕ_ТДУМБ_ЬСЗЕДНЕЗМАОР_ТУ
17. _ЕАЛЯРАНВЯАЧДА_ЕРПЕСАНВ_Ч
18. _И_ЕНТРЗИ_ОКЕВНОДЛЕША_ИМП
19. РОБДОЕВПС_МСХЬА_ _ИВПСНИОТ
20. ЕСДНОГТЕАНН_НЕОВМР_ЕУНПТЕ
21. _ЙЕСТОВО_НИИНЛАЕТИЖДСОПВ_
22. НДИАЕОЫЛППНЕ_ _НВЕАНГТ_ИЗЛА
23. П_БИРДЛЬНЕВ_ОП_ОПЗДЕВЫГЕА
24. МДООИТЕЬ_СМТ_НАДТЕСУБЕХНО
25. АИНАЛЖНОЛЕШФ_ЗИ_УАРОЬСНЕ_