УДК 519.72

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ

Королёва М.Н., Липницкий В.А.
УО «БНТУ», Минск, Беларусь, margo010172@rambler.ru;
УО «ВА РБ», Минск, Беларусь, valipnitski@yandex.ru

Современное информационное общество платит высочайшую цену за точную, надёжную и достоверную информацию. В корне подавляющего большинства катастроф техногенного характера обязательно найдутся дефекты информационного плана в системах управления – будь то космического или авиационного полёта или же какого-то технологического процесса. Серьёзный современный военный конфликт обязательно начинается с кибернетической атаки – атаки на информационные сети противника.

Главная проблема во всех системах обслуживания информационно коммуникационных потоков – обеспечение трёх названных выше качеств передаваемой информации. Решение проблемы достигается разработкой, созданием и развитием систем защиты информации, защиты от всякого рода помех и защиты от несанкционированного доступа к ней. Оба направления защиты информации имеют продолжительную историю, а их организация, в свою очередь, требует непременного применения информационных технологий.

Защита информации от помех наибольших успехов добилась для цифровых инфокоммуникационных систем (ИКС). Их прообразом явились телетайпы – первые информационные автоматы, успешно работавшие в почтовых системах связи в середине XX века.

Данные в телетайпах передавались в двоичной форме. Они были организованы в соответствии со строго сформированным протоколом. Данные в телетайпах передавались в так называемом ASCII-формате, то есть байтами – блоками по 8 бит. В каждом байте 7 бит были информационными и могли иметь произвольные значения. Восьмой бит был проверочным – всегда вычислялся, им был нуль или единица, причем таковым, что в целом в байте оказывалось чётное количество единиц, то есть двоичная сумма всех элементов байта равнялась нулю. Если приёмная часть телетайпа получала байт с нечетной суммой, то обязательно выполняемое им двоичное суммирование элементов байта давало «единицу». Это явно свидетельствовало о приобретении байтом ошибки в процессе его передачи. Автомат такой байт отбраковывал, на приёмное устройство телетайпа поступало требование о повторной передаче принятого байта – до получения байта с неизменной нулевой суммой.

Современные почтовые телекоммуникационные системы (ТКС), для которых скорость передачи информации не является приоритетным фактором, функционируют примерно также, только байты в них стали крупнее и проверочных разрядов в них стало существенно больше.

Следует признать, что существует определённая часть цифровых ТКС, для которых проблема борьбы в помехами решена полностью технологическим путём. Это волоконно-оптические ТКС. Для них статистика ложных бит идеальна: один ошибочный бит может появиться лишь в течение часа-двух непрерывного функционирования системы! Данные ТКС имеют практически два недостатка – они стационарны и дороги в денежном измерении.

Подвижные ИКС – системы сотовой связи, диспетчерские службы, авиасвязь, космическая связь, морское и военное дело, а также многие-многие иные, как правило, высокоскоростные – функционируют в условиях неизбежно загрязнённых каналов передачи информации разного рода шумами, излучениями, интерференциями и

прочими помехами. Высокая скорость их работы обуславливает необходимость совершенно иных подходов, иной организации достоверной доставки надежной информации, обязательной синхронной борьбы с помехами, что разом обеспечивается применением помехоустойчивого кодирования [1 – 3].

Гениальная идея Клода Шеннона о введении избыточности в передаваемую информацию для успешной борьбы с помехами [1, 2] была подтверждена первыми помехоустойчивыми кодами – кодами Хемминга [4]. Это совершенные линейные коды [2, 5, 6], не только обнаруживающие ошибки (как телетайпы), но и исправляющие их. Правда, системы, исправляющие лишь одну-единственную ошибку на передаваемый блок сообщения.

Преобразование информации от источника – текста, речи, музыки, изображений и т.п. – в цифровую информацию проводится достаточно стандартными методами, требует известных информационных технологий, а потому остаётся в тени наших рассмотрений. Этот блок преобразований осуществляет кодер источника передаваемой информации. Также мы не рассматриваем специально обратную процедуру – стандартное преобразование восстановленной цифровой информации в исходный, присущий источнику, или иной, нужный получателю вид, что осуществляется в декодере источника.

Организация ТКС на основе того или иного кода Хемминга требует хорошей математической подготовки специалистов, ряда предварительных действий и протокольных мероприятий, создания специальных аппаратных и/или компьютерных устройств и программ по реализации соответствующих информационных технологий.

Если применяется двоичный код Хемминга, то передача информации организуется блоками по $n = 2^m - 1$ двоичных символов. Из них $k = 2^m - 1 - m$ являются информационными, а m – проверочными. На передающем конце ТКС должна быть зафиксирована порождающая $(k \times n)$ – матрица G, а на приёмном конце – проверочная $(m \times n)$ – матрица H. Столбцы проверочной матрицы кода Хемминга представляют собой последовательную m – разрядную двоичную запись чисел 1, 2,…, $2^m - 1$. Алгоритм построения проверочной матрицы H достаточно очевиден. Строки порождающей матрицы G составляются из координат базисных векторов ядра проверочной матрицы H. Поэтому для построения порождающей матрицы необходимо решить систему линейных алгебраических уравнений:

$$H \cdot \overline{x}^T = \overline{0}^T .$$

Передаваемая информация в коде Хемминга представляет собой произвольные векторы $\overline{i}$ из двоичного k – мерного векторного пространства. На передающем конце они предварительно кодируются по формуле:

$$\overline{c} = \overline{i} \cdot G . \qquad (1)$$

Передаются в ТКС кодовые слова $\overline{c}$, образующие k – мерное линейное подпространство в n – мерном двоичном пространстве. Собственно, это подпространство и называется линейным (n, k) – кодом Хемминга и часто носит специальное обозначение C_χ. Известно, что все $\overline{c} \in C_\chi$ имеют вес $\omega \geq 3$.

Если кодовое слово не подверглось искажениям в канале передачи информации, то на приемном конце ТКС это будет достаточно быстро установлено, поскольку там каждое принятое сообщение $\overline{x}$ проверяется вычислением произведения:

$$S = H \cdot \overline{x}^T . \qquad (2)$$

Очевидно, величина S в формуле (2) является m – мерным вектором. Если $\overline{x} = \overline{c}$, то есть является кодовым словом, то $S = \overline{0}$ по определению ядра матрицы.

Если же $\overline{x} = \overline{c} + \overline{e}$ для вектора-ошибки $\overline{e}$, то $S = H \cdot (\overline{c} + \overline{e})^T = H \cdot \overline{e}^T \neq \overline{0}$ при условии, что $\overline{e}$ не является кодовым словом.

При внимательном рассмотрении выясняется, что столбцы проверочной матрицы кода Хемминга представляют собой все возможные ненулевые векторы двоичного m – мерного пространства. Это означает, что ненулевой вектор S совпадает с одним из столбцов матрицы H. Если $\overline{e}$ – вектор весом 1 – одиночная ошибка на каком-то j – м месте, то, как несложно видеть, вектор S будет совпадать именно с j – м столбцом матрицы H. Таким образом, очевиден идеальный способ коррекции одиночных ошибок кодом Хемминга: в принятом сообщении $\overline{x}$ надо инвертировать j – ю координату, определяемую вектором S.

После восстановления истинного переданного кодового слова $\overline{c}$ приёмное устройство должно из вектора $\overline{c}$ восстановить информационный вектор $\overline{i}$. Линейная алгебра гарантирует однозначность этого восстановления, поскольку формула (1) преобразуется в систему линейных алгебраических уравнений относительно неизвестных координат информационного вектора $\overline{i}$. Имеется большой выбор вариантов для алгоритмического и/или аппаратного решения этой системы уравнений. Самый простой получается, если изначально матрицу H выбрать систематической, то есть вида $H = (\mathrm{E}_k \,|\, K)$, где E_k – единичная матрица порядка k. Тогда у каждого кодового слова проверочными являются последние m координат. Их отбрасывание превращает вектор $\overline{c}$ в информационный вектор $\overline{i}$.

В целом, ТКС на основе кода Хемминга, как и на основе любого иного линейного кода, превращается в сложную многоэтапную информационную систему, которую обобщенно принято представлять в виде следующей схемы (рис 1.). Здесь система телекоммуникационной связи соединяет источник данных с получателем данных посредством канала связи.

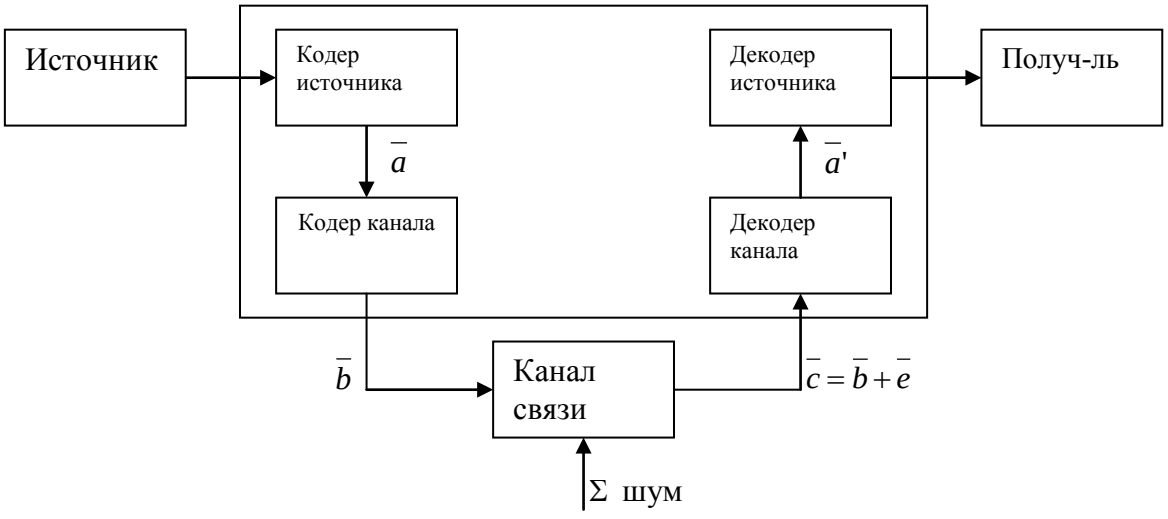

Рисунок1 - Обобщенная схема цифровой системы связи

Данные, поступающие в систему цифровой связи от источника данных, обрабатываются кодером источника – разбиваются на компактные блоки, которые преобразуются в стандартные последовательности символов – кодовые слова источника – информационные векторы $\overline{a} = \overline{i}$. Кодер канала преобразует их в кодовые слова канала $\overline{b} = \overline{c}$ – обычно более длинную последовательность символов, содержащую в себе некоторую избыточность. Каждый его символ может быть представлен битом или группой битов информации.

Кодовое слово канала передается по каналу связи (преобразованное модулятором канала связи в соответствующую последовательность аналоговых электромагнитных или иных сигналов, записанное в ЗУ для хранения и т.п.). В канале связи возможны разного рода шумы, искажения, интерференции (как естественного, так и наведенного происхождения). Поэтому на выходе из канала связи часто появляется информация отличающаяся от передаваемой, и демодулятор может выдать кодовое слово с ошибками. Декодер канала связи проверяет сообщение на истинность и при наличии ошибок исправляет их, используя избыточность в кодовом слове.

В настоящее время известно много различных кодов, одни из них контролируют любые ошибки кратности $k \geq 1$, другие рассчитаны на коррекцию специальных классов ошибок. Наиболее широкое применение получили линейные коды и их подкласс – циклические коды.

К сегодняшнему дню разработан достаточно широкий спектр помехоустойчивых кодов, позволяющих корректировать многократные ошибки. Правда, их построение требует применения новых разделов математики – теории групп и теории автоморфизмов групп, теории полей и полей Галуа [7 – 9]. При использовании различных кодов общая схема цифровой ИКС внешне остаётся неизменной. Изменяется внутреннее содержание блоков этой схемы. При этом наибольшим и радикальным изменениям подвергается третий блок схемы на рис.1 – блок декодера канала. Ведь для декодирования многократных ошибок построены самые разнообразные реализующие алгоритмы. Столбцы проверочной матрицы любого линейного кода по-прежнему характеризуют и однозначно определяют лишь ошибки весом 1.

Основные алгоритмические отличия для разных кодов обнаруживаются именно на этапе декодирования. При всем многообразии помехоустойчивых кодов практическую реализацию в конкретных действующих кодеках получил лишь небольшой их спектр. Это объясняется отсутствием хороших алгоритмов коррекции ошибок как с точки зрения быстродействия этих алгоритмов, так и с точки зрения аппаратурной сложности.

Так, коррекция двойных ошибок в сотовых системах связи реализуется кодами Боуза-Чоудхури-Хоквингема (БЧХ-кодами) и сводится к решению квадратных уравнений над полями Галуа $GF(2^m)$ из 2^m элементов – крайне плохо алгоритмизируемой процедуре [3, 6, 10]. Для преодоления этой сложности, а также преодоления «проблемы селектора» [11], возникающей при коррекции многократных ошибок из-за переборной сложности нахождения $\overline{a} = \overline{i}$ по вычисленному вектору S, белорусской школой помехоустойчивого кодирования разработана теория норм синдромов (ТНС) [12, 13]. ТНС допускает реализацию декодеров на высокоскоростных интегральных схемах, подобных нейронным сетям.

Современная защита информации от несанкционированного доступа начинает свою историю с публикации Уитфрилда Диффи и Мартина Хелмана [14] 1976 года. Здесь были высказаны три важные, основополагающие идеи: декларировано существование односторонних функций, возможность применения открытых ключей, протокол публичного обмена ключами. Три внешне простые мысли в корне

преобразили облик современной криптографии, сделали ее массовым явлением, проникшим во все сферы жизнедеятельности общества.

На основе идей Диффи У. и Хелмана М. возникли криптографическая система Ривеста Р, Шамира А. и Адлемана Л. (криптосистема RSA), а затем и криптосистема Рабина. Они отличались использованием открытых ключей и опорой на вычисления в кольцах классов вычетов Z / NZ с большим модулем $N = p \cdot q$ – произведением двух больших простых нечётных чисел. Криптографическая стойкость данных систем держалась на экспоненциальной сложности факторизации числа N на простые множители – исторически первой односторонней функции.

Данные криптосистемы привнесли в защиту информации новейшие достижения теории чисел, послужили толчком дальнейшего развития систем компьютерной алгебры, в частности, математического пакета «Математика». Дальнейшему совершенствованию подверглись методы вычислений с большими числами. Легальные пользователи этих криптосистем стали активно применять в своих вычислениях китайскую теорему об остатках в современном её звучании: «кольцо Z / NZ изоморфно прямому произведению $Z / pZ \times Z / qZ$ полей классов вычетов по простым модулям». Такой подход, по-существу, уполовинивал разрядность применяемых в криптографических вычислениях величин по сравнению с действиями хакеров или неумелых пользователей (от 120 – 150 десятичных разрядов до 50 – 75).

Хакерские атаки на эти криптосистемы указали на необходимость аккуратного применения ключей, но в целом создали ощущение, что только квантовый компьютер станет радикальным средством взлома данных криптосистем. Тем не менее, названные криптографические системы оказались весьма вязкими и медленными в практическом применении, что стимулировало поиск альтернативных криптографических систем.

Первой из них оказалась криптосистема Эль Гамаля. Она опирается на вычислениях больших степеней в поле классов вычетов Z / PZ для большого простого числа P – до 300 десятичных знаков в реальных криптосистемах. Криптографическая стойкость данной системы базируется на другой односторонней функции – вычислении дискретного логарифма, то есть нахождении целой степени x, такой, что $a^x = b$ для заданных a и b из Z / PZ. Криптосистема Эль Гамаля завоевала симпатии пользователей со всего мира. На ее основе были созданы криптографические стандарты различных стран мира, включая Россию и Республику Беларусь.

Через 20 лет эксплуатации криптосистемы Эль Гамаля и её аналогов выяснилось, что ещё в 1962 году в одной из секретных лабораторий был разработан достаточно эффективный алгоритм «baby step giant step» вычисления дискретного логарифма. Этот факт существенно подорвал доверие к этой весьма элегантной и красивой криптографической системе, реализация этапов которой требует относительно простых информационных технологий.

В наши дни практическую апробацию проходят различные аналоги криптосистемы Эль Гамаля – ЕКСТР-криптосистемы, эллиптическая криптография, различные теоретико-групповые конструкции. Вместе со стандартами шифрования DES и AES – они практически не оставили в тени ни одного новейшего результата теории чисел, теории алгоритмов, современной алгебры и алгебраической геометрии [15 - 17].

Реализация каждого из этапов шифрования, дешифрования, взлома названных криптографических систем требует своих подходов, своих индивидуальных алгоритмов и своих информационных технологий, как аппаратурных, так и мощных компьютерных средств.

Литература

1. Шеннон К. Работы по теории информации и кибернетике. – М.: ИЛ, 1963. – 332 с.
2. Shannon C.E. A mathematical theory of communication. Part I, II // Bell. Syst. Tech. J. – 1948. – Vol. 27. – P. 379–423; P. 623–656.
3. Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. Теория кодов, исправляющих ошибки: Пер. с англ. – М.: Связь, 1979. – 744 с.
4. Hamming R.W. Error detecting and error correcting codes. – Bell Syst. Tech. J., 1950. – P. 147 – 160.
5. Галлагер Р. Теория информации и надежная связь. – М.: Советское радио, 1974. – 720 с.
6. Блейхут Р. Теория и практика кодов, контролирующих ошибки. – М.: Мир, 1986. – 576 с.
7. Артин Э. Геометрическая алгебра. – М.: Наука, 1969. – 284 с.
8. Лиддл Р., Нидеррайтер Г. Конечные поля. Т. 1, 2. - М.: Мир, 1988. – 822 с.
9. Липницкий В.А. Современная прикладная алгебра. Математические основы защиты информации от помех и несанкционированного доступа. – Мн.: БГУИР, 2005. – 88 с.; 2-е издание: Мн.: БГУИР, 2006. – 88 с.
10. Вернер М. Основы кодирования. – Учебник для ВУЗов. - М.: Техносфера, 2006. – 288 с.
11. Колесник В.Д., Мирончиков Е.Т. Декодирование циклических кодов. – М.: Связь, 1968. – 251 с.
12. Конопелько В.К., Липницкий В.А. и др. Прикладная теория кодирования. Т. 1 – 2. – Мн.: БГУИР, 2004. – 688 с.
13. Липницкий В.А., Конопелько В.К., Норменное декодирование помехоустойчивых кодов и алгебраические уравнения. – Мн.: Издательский центр БГУ, 2007. – 240 с.
14. Diffie W. and Hellman M.E. New Directions in Cryptography. – IEEE Trans. Inf. Theory, 1976. – P. 644 – 654.
15. Василенко, О. Н. Теоретико-числовые алгоритмы в криптографии / О. Н. Василенко. – М. : МЦНМО, 2003. – 326 с.
16. Зензин, О. С. Стандарт криптографической защиты AES. Конечные поля / О. С. Зензин, М. А. Иванов. – М. : Кудриц-Образ, 2002. – 168 с.
17. Смарт Н. Криптография. – М.: Техносфера, 2005. – 528 с.
18. Харин Ю.С. и др. Криптология: учебник. – Мн.: БГУ, 2013. – 512 с.