УДК 621.391.251

ЗАЩИТА ИНФОРМАЦИИ НА ОСНОВЕ СЖАТИЯ

А.А. Борисенко, профессор
Сумский государственный университет

Рассмотрены вопросы защиты информации от несанкционированного доступа на основе алгоритмов сжатия. Проведен качественный анализ эффективности такой защиты. Рассмотрен подход к защите информации на основе двух источников информации, один из которых представляет источник открытых, а другой - закрытых ключей.

Как методы сжатия, так и методы защиты информации широко применяются на практике, но совместно они используются довольно редко, так как считается, что защита информации при сжатии недостаточна. Однако даже такая нестойкая защита в ряде случаев достаточна для не слишком важной информации, теряющей свою ценность в течение нескольких часов или даже десятков минут, например, при оперативном управлении быстротекущими процессами. При этом малое время на шифрование и передачу информации дает значительное преимущество по сравнению с другими, более стойкими шифрами.

Любое сжатие информации - это есть пусть и небольшая, но все же ее защита от несанкционированного доступа, так как при сжатии реализуется функция преобразования $F : X \to Y$, где X представляет исходное множество сообщений, а Y - множество, в которое они преобразуются под воздействием алгоритма преобразования, представляющего собой алгоритм их сжатия.

При этом в каждом таком алгоритме имеются два ключа – открытый и закрытый (секретный).

Открытые ключи представляют собой сжатые сообщения, так как именно они несут основную информационную нагрузку для приемника информации. Чтобы их восстановить, нужно знать обратный алгоритм преобразования в исходные сообщения и, кроме того, надо иметь тайный ключ для данного сообщения.

Так, например, если использовать алгоритм кодирования Хаффмана для защиты информации, как это предлагается сделать, например, в [1], то, кроме этого алгоритма, надо знать распределение вероятностей на входе преобразователя. Обычно она для заданного класса сообщений определяется статистически и поэтому становится довольно быстро известной не только приемнику информации, а и лицу, которое пытается открыть этот код. Однако все же для определения этих вероятностей раскрывающему лицу надо потратить определенное время, за которое ценность переданной информации может уменьшиться, и не будет смысла ее вскрывать.

Обратим внимание, что в данном случае использовались известный алгоритм кодирования исходных символов и два ключа: первый – открытый, представляющий сжатые кодовые последовательности, и второй – закрытый, задающий вероятности генерирования тех или иных

символов исходного алфавита. По этим вероятностям затем строится оптимальный код, как, например, в табл. 1.

Таблица 1- Код Хаффмана для латинского алфавита

Буква	Код Хаффмана	Буква	Код Хаффмана
E	100	M	00011
T	001	U	00010
A	1111	G	000001
O	1110	Y	00001
N	1100	P	110101
R	1011	W	011101
I	1010	B	011100
S	0110	V	1101001
H	0101	K	110100011
D	11011	X	110100001
L	01111	J	110100000
F	01001	Q	1101000101
C	01000	Z	1101000100

С помощью этого кода затем формируются защищенные передаваемые сообщения [1]. Например, строка текста из 29 знаков

WENEEDMORESNOWFORBETTERSKIING

преобразуется в строку двоичных комбинаций:

01110110 01100100 10011011 00011111
01011100 01101100 11100111 01010011
11010110 1100100 00100110 01011011
01101000 11101010 10110000 001.

Данный код можно усложнить, образуя символы сообщений, собирая их в пакеты по два - три символа, и так до размера всего сообщения. Затем эти пакеты кодируются с помощью новых вероятностей, которые образуются как произведения исходных. При этом изменяются открытый и закрытый ключи. Последний из них также легко открывается, как и ранее. По нему и известному алгоритму сжатия открытый ключ преобразуется в исходное сообщение.

Однако существуют более сложные методы сжатия на основе кодов Хаффмана и Шеннона, предлагаемые автором [2], когда используются детерминированные ограничения. В дальнейшем они представляются для каждого сообщения как вероятностные и имеют свою особую форму, которую нельзя определить на основе статистических испытаний.

В этом случае приходим к двум особым источникам информации – комбинаторному и вероятностному.

Комбинаторный источник вырабатывает по существу номера сообщений, представляющие открытые ключи, а вероятностный источник – закрытые ключи, которые может знать, кроме источника, только приемник информации.

Эти ключи генерируются для каждого передаваемого сообщения и могут при необходимости быть зашифрованы с помощью любых стандартных шифров или тайно переданы с открытыми ключами как по общему каналу, так и по выделенному.

Так как открытые ключи генерируются комбинаторным источником информации, представляя по сути номера информационных сообщений, то их появление равновероятно. Это значит, что их нельзя восстановить, используя статический подход. Их число для двоичного канала равно 2^n, где n - длина открытого ключа (сжатого сообщения). Чтобы открыть

(восстановить) это сообщение, необходимо организовать последовательный перебор всех 2^n сообщений. При $n > 100$ это сложная, если вообще выполнимая, задача. Вероятность угадать случайным образом также ничтожно мала. Поэтому вероятность раскрытия такого сообщения без знания закрытого ключа небольшая. При этом, если ключ даже для какого-то сообщения будет найден, то для следующего он автоматически будет изменен. Для его определения снова понадобится время, которое может быть очень большим. Процедура получения новых ключей для новых сообщений близка к процедуре гаммирования. Это значит, что данный подход дает труднораскрываемую защиту на уровне лучших современных криптографических систем. При этом следует учесть, что происходит сжатие информации, а значит, уменьшается время ее передачи и, кроме того, шифруется только малая часть передаваемой информации, в среднем около 10%. Все эти достоинства позволяют говорить о высокой потенциальной эффективности методов засекречивания на основе особых комбинаторных методов сжатия информации и, следовательно, о необходимости проведения исследований в данном направлении. Причем, судя по всему, такие коды должны превзойти по стойкости и эффективности в целом существующие на сегодня методы защиты информации. К их достоинствам, кроме вышеперечисленных, относятся возможность эффективной аппаратной реализации и относительная простота алгоритмов кодирования.

Недостаток рассматриваемых методов защиты данных – это снижение помехоустойчивости при передаче информации за счет ее сжатия. Однако этот недостаток легко устраняется помехоустойчивым кодированием.

На сегодня существует ряд комбинаторных методов сжатия информации, среди которых разработанные автором - биномиальные, которые могут быть применены для эффективной защиты информации. Особенно интересны в этом направлении нумерационные методы сжатия и в целом нумерационное кодирование.

Важно также и то, что некоторые из методов сжатия способны защищать не только передаваемую информацию, а и обрабатываемую, то есть с их помощью можно осуществить передачу, хранение и обработку информации в едином цикле.

Таким образом, можно утверждать, что использование сжатия для защиты информации от несанкционированного доступа имеет ряд несомненных достоинств, среди которых повышение скорости и высокая эффективность защиты, которая значительно превосходит ряд используемых на практике систем сжатия.

SUMMARY

The questions of data defence from unauthorized access on basis of compression algorithms are under review in the paper. Qualitative analysis of efficiency of such defence is carried out. The considered method of data security uses two information sources, one of which is of a public keys source, but the other one – a secret keys source.

СПИСОК ЛИТЕРАТУРЫ

1. Хаффман Л.Дж. Современные методы защиты информации /Пер. с англ. - М.: Советское радио, 1980. - 262 с.
2. Борисенко А.А Комбинаторное сжатие информации //Вісник СумДУ. - 1996. - №1(5). – С. 79-83.

Поступила в редакцию 12 мая 2006 г.