УДК 681.3.06: 519.248.681

АНАЛИЗ СТОЙКОСТИ АЛГОРИТМА ГОСТ 28147-89 ПРИ ИСПОЛЬЗОВАНИИ ПОДСТАНОВОК СЛУЧАЙНОГО ТИПА

ГОРБЕНКО И.Д., ЛИСИЦКАЯ И.В., КОРЯК А.С.

В статье приведены результаты анализа статистической безопасности симметричного алгоритма криптопреобразования ГОСТ 28147-89 при использовании случайных таблиц подстановок предложенных в АО "ИИТ".

В [1,2] обсуждался вопрос применения в алгоритме ГОСТ 28147-89 подстановок случайного типа. Предложены критерии для отбора случайных подстановок и случайных таблиц подстановок, которые предлагалось использовать в этой процедуре шифрования в качестве долговременных ключей (*S*-блоков). В [1] также приведены аргументы в пользу развиваемой методики построения случайных таблиц подстановок, базирующиеся на простых физических и математических соображениях.

Изложим результы проведения более строгой криптографической проверки эффективности применения случайных таблиц подстановок в алгоритме ГОСТ 28147-89.

Как известно [3], статистические испытания являются единственной стратегией испытаний больших криптографических систем с секретными ключами, построенных в виде чередующихся слоёв блоков замен и перестановок, как это сделано в системах DES, ГОСТ. Это объясняется трудностью составления уравнений, связывающих входы и выходы таких систем, которые можно было бы решать строгими аналитическими методами. Поэтому статистические испытания стали главным инструментом наших исследований.

За основу взяты три показателя стойкости (статистической безопасности [3]), которые принято сейчас использовать при проверке многоцикловых процедур современных блочных систем шифрования:

1. Число циклов алгоритма, начиная с которого две криптограммы, полученные шифрованием двух отличающихся на один бит блоков данных (открытых текстов), становятся устойчиво независимыми (в том смысле, что при большем числе циклов они остаются независимыми). Другими словами, необходимо определить число циклов шифрования алгоритма, начиная с которого обеспечивается влияние любого (одного) входного бита на каждый выходной бит – лавинный эффект [4,5].

2. Число циклов шифрования, при котором один и тот же открытый текст, зашифрованный на ключах, отличающихся на один бит, порождает устойчиво независимые (некоррелированные) криптограммы.

3. Коэффициент сжатия шифрованного текста при применении процедуры архивирования Лемпела-Зива, характеризующий степень его случайности.

При обосновании эффективности случайных подстановок первой стала задача — изучить влияние изменений входных бит шифруемого блока на выходные биты для каждого цикла процедуры шифрования. Сначала были рассмотрены в этом отношении свойства и характеристики таблицы подстановок, представленной в [4] в качестве долговременного ключа ГОСТа. Результаты статистических экспериментов для этого случая поданы в табл. 1, где m_W – математическое ожидание числа $W(D_k)$ единичных (ненулевых) бит в булевой побитной сумме D_k (сумме по модулю 2) пары шифртекстов на k-м цикле алгоритма шифрования, открытые тексты которых отличаются одним битом; σ_W^2 – дисперсия числа единичных бит для этой же побитной суммы.

Для того чтобы выделить в чистом виде влияние на процедуру преобразования именно долговременных ключей (*S*-блоков), в качестве сеансового ключа в первых двух колонках табл. 1 использовался нулевой вектор, $K_0 = K_1 = \ldots = K_7 = 0$ (см. описание ГОСТа в [5]), что обозначено в табл.1 вектором $\overline{K} = 0$. В правых двух колонках табл.1 обозначение $\overline{K} \neq 0$ соответствует использованию в качестве сеансового ключа произвольного ненулевого вектора. Здесь отмечены позиции в парах открытых текстов с отличающимися битами.

Возникает вопрос, каким образом определять номер цикла, после которого можно считать, что изменение входного бита влияет сразу на все выходные биты цикла?

С одной стороны, математическим свидетельством влияния изменения бита в открытом тексте одновременно на все выходные биты цикла является ортогональность (некоррелированность) пары шифрованных текстов, соответствующих отличающимся на один бит открытым текстам, т.е. значение m_W = 32 (для ГОСТа). Действительно, если рассматривать чисто случайный 64-битный блок (состоящий из независимых и равновероятных двоичных символов), то закон распределения числа ненулевых (нулевых) бит в таких блоках будет биномиальным с параметрами

$$m_W = np_0 = 64 \cdot \frac{1}{2} = 32,$$

$$\sigma_W^2 = np_0(1-p_0) = 64 \cdot \left(\frac{1}{2}\right)^2 = 16.$$

При значении $np_0(1-p_0) \geq 10$ биномиальное распределение с высокой степенью приближения аппроксимируется нормальным законом распределения вероятностей (формула Муавра-Лапласа [6]).

С другой стороны, формируемые оценки m_W являются случайными и редко равны точно 32. Однозначный ответ на поставленный выше вопрос можно получить, если воспользоваться методом доверительных интервалов, т.е. методом математической статистики, специально предназначенным для построения множества приближенных значений неизвестных параметров вероятностных распределений. В кратком изложении сущность этого метода состоит в следующем.

Пусть $X_1, X_2, \ldots, X_n$, $n \geq 2$, – независимые случайные величины, подчиняющиеся одному и тому же нормальному закону с неизвестными пара-

Таблица 1

метрами $EX_i = \theta_1$ и $DX_i = \theta_2$, причём требуется построить интервальную оценку $u(\theta) = \theta_1$. Пусть

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \,;\; s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X - \overline{X})^2 .$$

Поскольку случайная величина $T = \sqrt{n}\,(\overline{X} - \theta)/s$ подчиняется распределению Стьюдента с $n-1$ степенями свободы и не зависит от неизвестных параметров q_1 и q_2 ($|\theta_1| \le \infty, \theta_2 > 0$), то при любом положительном t вероятность события

$$\left\{ \overline{X} - \frac{t \cdot s}{\sqrt{n}} < \theta_1 < \overline{X} + \frac{t \cdot s}{\sqrt{n}} \right\}$$

зависит лишь от t. Если указанный интервал принять за оценку c для q_1, ему будет соответствовать доверительная вероятность

$$P_c(\theta_1, \theta_2) = P\{|T| < t\} = 1 - \alpha,$$

не зависящая от $\theta = \{\theta_1, \theta_2\}$. Такую интервальную оценку называют доверительным интервалом, а его концевые точки – доверительными пределами.

В соответствии с изложенным методом, задаваясь доверительной вероятностью $P_c(\theta_1, \theta_2) = 0{,}999 \;\rightarrow (\alpha = 0{,}001)$, на основании таблицы распределения Стьюдента для заданных значений α и $n = 1024$ (в наших опытах) получаем $t = 3{,}291$ и, следовательно,

$$\frac{t \cdot s}{\sqrt{n}} = \frac{3{,}291 \cdot 4}{\sqrt{1023}} = 0{,}4.$$

Это значит, что можно считать попавшими в доверительный интервал все значения m_w, удовлетворяющие следующим условиям:

$$32 - 0{,}4 \le m_w \le 32 + 0{,}4.$$

В результате отсчёт числа циклов для алгоритма ГОСТ, при котором считается, что изменение входного бита практически не влияет на все выходные биты, можно вести по моменту "накрытия" случайного интервала $(\overline{X} - 0{,}4,\ \overline{X} + 0{,}4)$, выступающего в качестве доверительной оценки математического ожидания $\overline{X}$ с мерой надёжности $1 - \alpha = 0{,}999$, значение $m_w = 32$. Оно соответствует полной сбалансированности (ортогональности) шифрованных текстов сообщений, отличающихся для рассматриваемого цикла одним битом.

Имея теперь методику оценки результатов, перейдём к их анализу. Как следует из табл. 1, при использовании в алгоритме ГОСТ его "родной" таблицы подстановок и ненулевого сеансового ключа ($\overline{K} \neq 0$), для обеспечения зависимости всех выходных бит от любого входного бита требуется вхождения в алгоритм на 8-9 циклов. Результат, приведенный для 1-го и 64-го битов, практически сохраняется и для других бит.

При нулевом сеансовом ключе $\overline{K} = 0$ появляется ещё один дополнительный "этаж". В целом получается, что характеристики перемешивания в статистическом смысле мало зависят от сеансового ключа $\overline{K}$ (цена – один цикл).

Аналогичные статистические эксперименты были проведены и при использовании в качестве долговременных ключей в алгоритме ГОСТ таблицы из подстановок вырожденного типа (составленной из тождественных подстановок) и таблиц подстановок, отобранных (сформированных) по предложенным в работе [1] критериям. Для наглядности и компакт-

ности эти результаты представлены на рис. 1 в графическом изображении, здесь же приведены и результаты для ГОСТ-подстановки из табл.1.

На рис. 1 показана зависимость математического ожидания числа ненулевых бит m_W в побитовой булевой сумме двух шифрованных текстов, исходные тексты которых отличаются на один бит от числа циклов алгоритма. Глубина вхождения в алгоритм для таблицы, составленной из тождественных подстановок, при ненулевом сеансовом ключе увеличилась до 14 циклов. К тому же, здесь оказывается

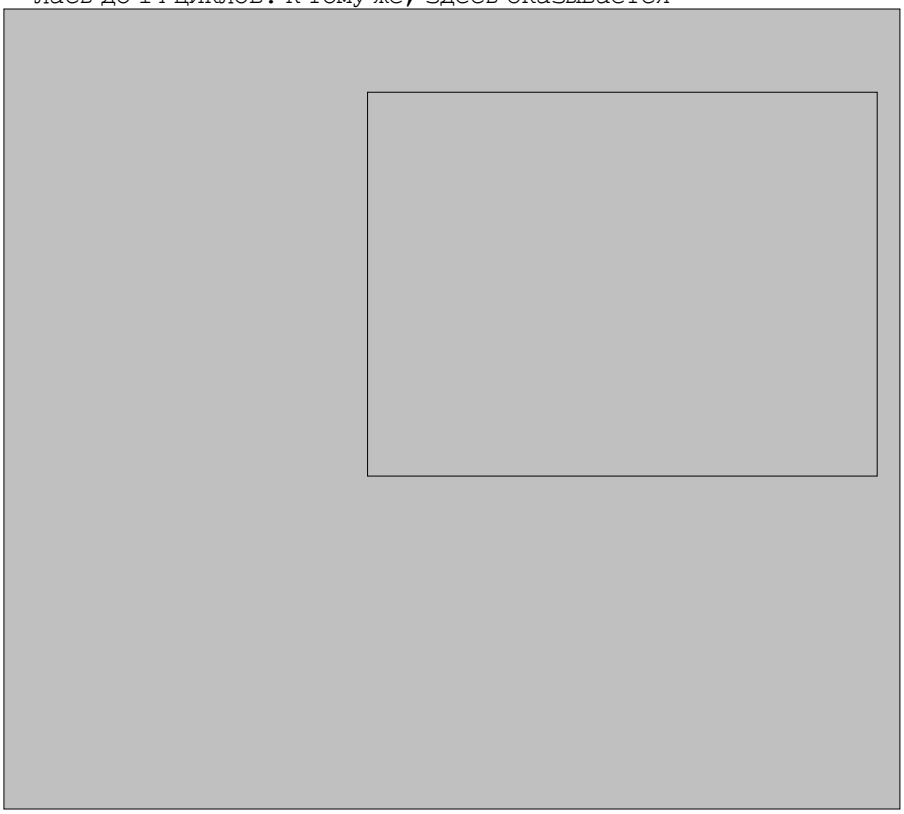


Рис. 1. Зависимость математического ожидания при изменении одного бита исходного текста от числа циклов

существенным применение ненулевого сеансового ключа ($\overline{K} \neq 0$), так как при $\overline{K} = 0$ получается вырожденный результат. Результаты экспериментов при изменении других бит и использовании других сеансовых ключей оказались статистически эквивалентными представленным на рис. 1. Здесь же даны результаты испытаний для одной из таблиц подстановок, прошедших предложенную в работе [1] методику проверки. Общим является то, что при изменении долговременного ключа (рассматривались в том числе и таблицы, составленные из противоречивых и одноцикловых подстановок, прошедших проверки на случайность) характеристики случайных таблиц подстановок в целом оказываются очень близкими к свойствам таблиц из представленных примеров (для ГОСТ-овской и сгенерированной по правилам [1]).

Очередным этапом статистических испытаний стала оценка числа циклов алгоритма, при котором наступает статистическая независимость выходных текстов в процессе шифрования сообщений с помощью ключей $\overline{K}$, отличающихся одним битом.

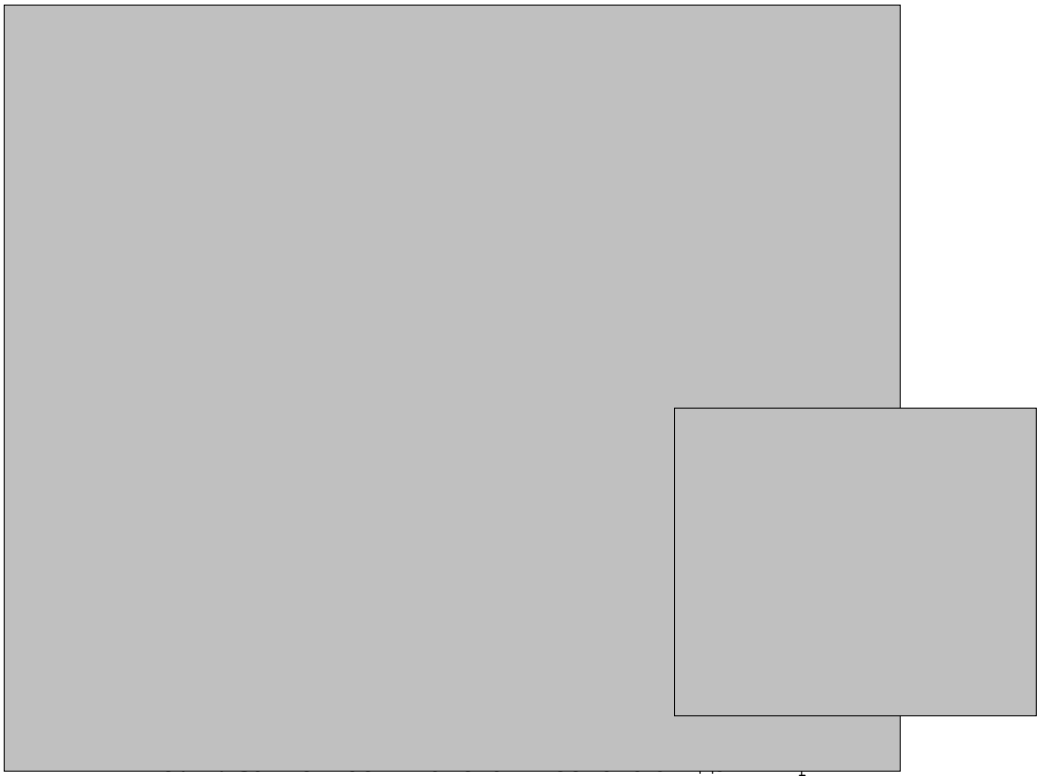
Простые рассуждения показывают, что в соответствии с полученными выше результатами и процедурами преобразований в каждом цикле [8] в рассматриваемом случае необходимо ожидать, что "глубина" T_w вхождения в алгоритм будет зависеть от позиции измененного бита в матрице КЗУ.

Если изменяется бит в строке, которая используется уже на первом цикле, следует ожидать значения T_w близкого к 9 (воздействие измененного бита ключа эквивалентно изменению одного бита сообщения). Если же изменяется бит в последней строке, которая используется в алгоритме, начиная с 8-го цикла, ожидаемым значением T_w будет величина 9 + 7 = 16. Естественно, что во всех циклах, предшествующих циклу с измененным значением ключевого бита, результат побитного суммирования шифртекстов для одного и того же сообщения будет давать значения $m_W = 0, \quad \sigma_W^2 = 0$. Результаты экспериментов, представленные на рис. 2, 3, полностью подтвердили эти предположения. Здесь показана зависимость математического ожидания числа ненулевых бит m_W в побитовой булевой сумме двух текстов, зашифрованных с помощью сеансовых ключей, отличающихся одним битом от числа циклов алгоритма

Общим результатом можно считать увеличение значения T_w до 21 (для тождественной подстановки), т.е. до конца процедуры еще остается 11 циклов со сбалансированным перемешиванием.

На рис. 4 представлены результаты статистических испытаний для случая, когда переменными параметрами брались уже символы долговременного ключа (ненулевые строки таблицы подстановок $S_{m,n}$). Ожидалось, что при изменении долговременного ключа он будет влиять на процедуру шифрования многократно (вносить новые изменения на каждом цикле), т.е. лавинный эффект размножения ошибок будет усиливаться. Это, однако, оказалось верным только при существенных изменениях долговременного ключа (кривые 2,3 на рис. 4). В то же время при небольших изменениях в таблице $S_{m,n}$ (при перестановке двух, трех, четырех символов) оказалось, что не удается дойти до заданного уровня и при всех 32 циклах алгоритма. Более того, как показывает анализ, в этом случае с большой вероятностью возможны ситуации, когда получается значение различия шифрованных текстов, равное нулю на всех циклах преобразования (искаженный ключ приводит при шифровании и расшифровании к тому же результату, что и действительный).

Такие ситуации возможны, когда в процессе шифрования (дешифрования) не используются искаженные переходы таблицы подстановок. Например, если в какой-то из строк переставлены местами два символа, то вероятность события, заключающе-



изменении одного бита строки сеансового ключа и тождественной таблицы подстановок от числа циклов

– в разных строках:

$$\left(\frac{14}{16}\right)^{64} = 2 \cdot 10^{-4}.$$

Транспозиция в каждой из строк:

$$\left(\frac{14}{16}\right)^{32 \cdot 8} = 1{,}3 \cdot 10^{-13}.$$

Для того чтобы сделать вероятность такой вырожденной ситуации меньше чем 2^{64}, необходимо иметь в таблице более трех искаженных переходов в каждой из строк:

$$\left(\frac{13}{16}\right)^{32 \cdot 8} = 0{,}82 \cdot 10^{23}.$$

Можно сделать вывод, что таблица, составленная из подстановок, выбранных наугад, с большой вероятностью будет обеспечивать высокие характеристики защиты информации. Однако в этом случае существуют хоть и маловероятные, но неблагоприятные ситуации (вырожденные ключи), для которых необходимые характеристики стойкости алгоритма не гарантируются.

Была также проверена степень сжатия текстов, зашифрованных на ключах – подстановках случайного типа, с помощью процедуры Лемпела-Зива

гося в том, что на данном цикле при прохождении искаженной строки подстановок она не повлияет на результат, будет равна вероятности выпадения (появления) на входе этого *S*-блока (задаваемого строкой с искаженными элементами) чисел, не дающих искаженного результата. Эта вероятность равна 14/16 (14 чисел из 16 не дают искажения). Для получения нулевого различия в шифрованных текстах необходимо, чтобы эта ситуация состоялась на каждом из 32 циклов, что дает результирующую вероятность (в предположении независимости рассматриваемых событий), равную

$\left(\frac{14}{16}\right)^{32} = 0{,}014$, т.е. в 14-ти случаях из 1000 искаженная подстановка будет шифровать текст точно так же, как и не искаженная. Приведем примеры просчетов и для некоторых других характерных случаев.

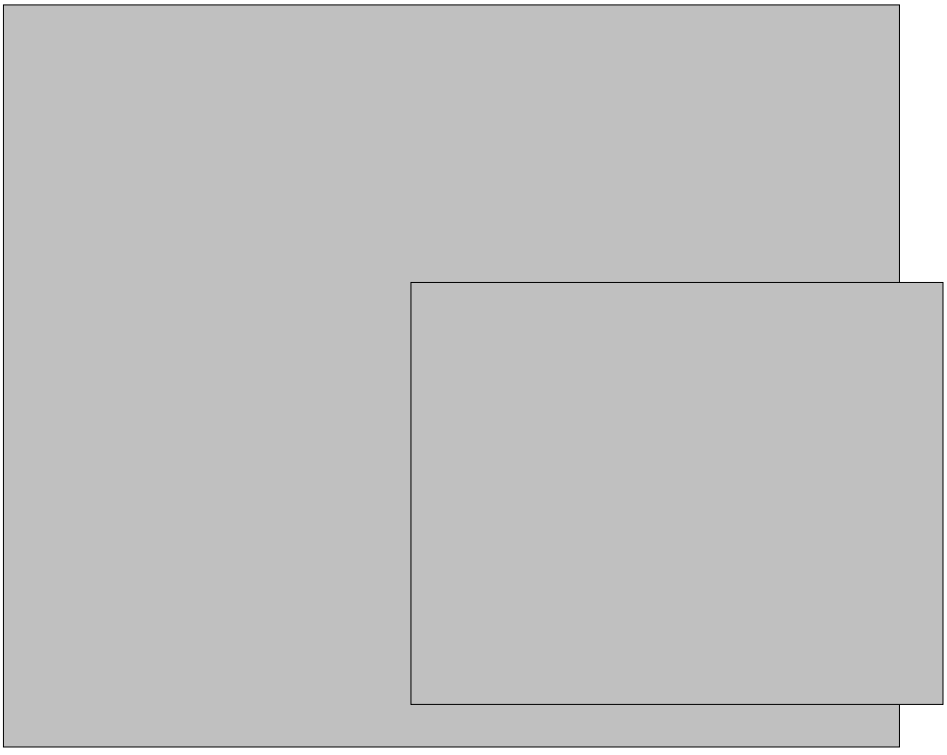


Рис. 3. Зависимость математического ожидания при изменении одного бита строки сеансового ключа от числа циклов

Две транспозиции различных элементов уже дают результаты:

– в одной строке:

$$\left(\frac{12}{16}\right)^{32} = 10^{-3};$$

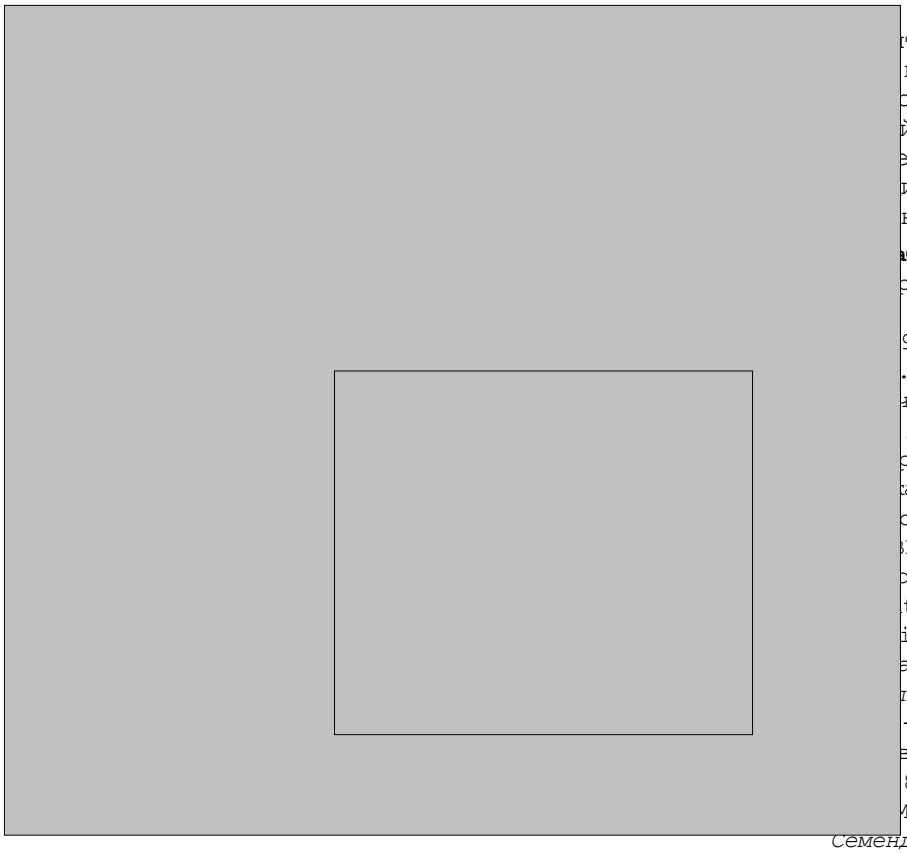


Рис. 4. Зависимость математического ожидания при изменении долговременной таблицы по ГОСТ и $(\overset{\rho}{K} \neq 0)$ от числа циклов

(проверка меры случайности текстов [3]). Испытывались три варианта текстов: exe-файл, как пример очень избыточного (с повторениями) текста, word-файл, как текст усредненного типа, и обыкновенный текстовый файл. Результаты этой проверки иллюстрирует табл. 2.

Как видно из представленных результатов, во всех случаях, кроме ситуации с шифрованием exe-файла, даже при шифровании на одних и тех же ключах (сеансовом и долговременном) обеспечивается сжатие шифрованного текста менее чем на 10% [3]. Результат с exe-файлом свидетельствует лишь о том, что в этом случае имеется значительная часть повторяющихся (одинаковых) сообщений (при шифровании exe-файла на разных сеансовых ключах он уже не сжимается).

Таблица 2

итоге можно сделать вывод, что по приведенным выше показателям ости (статистической безопасности) йные таблицы подстановок, постро- с помощью предлагаемой в [1] ики, не уступают свойствам долго- нного ключа, приведенного в [4].

атура: 1. *Горбенко И.В., Лисицкая И.В.* ории отбора случайных таблиц подста- для алгоритма шифрования по ГОСТ 9 // Радиотехника. 1997. Вып. 103. С.121- . *Горбенко И.D., Лисицкая И.В.* К оценке еских характеристик таблиц подста- для алгоритма криптографического разования по ГОСТ 2847-89 // Радио- а. 1997. Вып 104. С. 97. 3. *Жильников В.* ография от компьютера до папируса. F, 1996. 336с. 4. *Sheier B.* Applied ography. Second Edition: protocols, thms, and source code in C. Published by iley & Sons. Inc. New York: Chichester ane Toronto Singapore, 1996. 158 p. ь *А.М.* Алгоритм шифрования по ГОСТ -89 и способы применения блочных в // Безопасность информации. 1995. 8-11. 6. *Вентцель Е.С.* Теория вероятно- М.: Наука, 1964. 564 с. 7. *Бронштейн И.Н. Семендяев К.А.* Справочник по математике для инженеров и учащихся ВТУЗов. М.: Наука, 1980. 976 с. 8. ГОСТ 29147-89. Системы обработки информации. Защита криптографическая. Алгоритмы криптографического преобразования. Введ. 01.01.89. М.: Изд-во стандартов. 1989. 78 с.

Поступила в редколлегию 12.03.98

Горбенко Иван Дмитриевич, д-р техн. наук., профессор, проректор по научной работе ХТУРЭ. Научные интересы: защита информации в компьютерных системах и сетях. Адрес: 310726, Украина, Харьков, пр.Ленина 14, тел. 30-24-50, 37-56-39.

Лисицкая Ирина Викторовна, ассистент кафедры прикладной математики ХВУ. Научные интересы: защита информации. Адрес: 310000, Украина, Харьков, пр. Ленина, 82, кв. 32, тел. 32-08-38.

Коряк Алексей Сергеевич, студент ХТУРЭ. Научные интересы: компьютеры и программирование. Адрес: 310000, Украина, Харьков, пр. Гагарина, 48, кв. 24, тел. 27-18-80.