

## **О стойкости ГОСТ 28147–89**

Отечественный стандарт шифрования ГОСТ 28147–89 де-факто является одним из лучших в мире шифров в части скорости снижения стойкости. Если многочисленные зарубежные шифры (DES, FEAL, RC5, RIJNDAEL и др.) характеризуются появлением новых методов криптоанализа, снижающих их стойкость все сильнее, то для нашего стандарта не известны методы анализа, снижающие стойкость по сравнению с перебором.<sup>1</sup>

В 2001 г. А. Ростовцевым и Е. Маховенко были предложены новые методы криптоанализа, не требующие большого объема открытых и зашифрованных текстов и при этом более эффективные чем популярные линейный и дифференциальный методы криптоанализа [1]. Предлагаются результаты проверки метода рационального продолжения применительно к задаче вскрытия ключа ГОСТ 28147–89 на объеме статистики  $10^5$ . Показано, что разреженные ключи являются эффективно распознаваемыми и поэтому могут считаться слабыми. Ключи, содержащие одинаковое число нулей и единиц в каждом слове, характеризуются тем, что частота совпадения оценки бита ключа с истинным значением равна 0,59, что равносильно снижению шенноновской энтропии на один разряд ключа с 1 до 0,974.

*Rostovtsev A.G., Makhovenko E.B., Filippov A.S., Chechulin A.A.,  
SPbSPU*

## **On the strength of GOST 28147–89**

Russian encryption standard GOST 28147–89 de facto is one of the best ciphers in the world. If many foreign ciphers (DES, FEAL, RC5, RIJNDAEL, etc.) can be characterized by appearing new attacks, due to which cipher strength decreases more and more, there are no known cryptanalytic attacks for GOST 28147–89 that allow to decrease its strength with respect to enumeration.

Recently A. Rostovtsev and E. Makhovenko have suggested new cryptanalysis methods, that do not take large number of plaintext / ciphertext pairs and are more effective than popular differential and linear methods [1]. These methods were applied for GOST 28147–89 for statistic amount  $10^5$ . It is shown that sparse keys can be effectively recognized and so they are weak. If numbers of units and zeroes in each key word are the same, then frequency of event that key bit estimation equals to true key bit is 0.59. Hence key bit entropy is 0.974.

### **1. Методы анализа блочных шифров**

В настоящее время для обеспечения конфиденциальности обычно используются симметричные итерированные блочные шифры. Безопасность таких шифров основана на задаче вскрытия ключа по известным или подобранным открытым и зашифрованным текстам. Сложность этой задачи обусловлена тем, что трудно определить, насколько тестируемый ключ близок к истинному, так как небольшое изменение ключа или открытого текста вызывает значительное (в среднем 50%) изменение разрядов шифрограммы. Иначе говоря, трудно задать вычислимую функцию (метрику), показывающую «расстояние» между тестируемым и истинным ключом.

Для задания такой функции используют показатели, характерные для текстов и ключей «в среднем». При этом требуется большой (обычно практически недостижимый) объем статистики открытых и зашифрованных текстов, причем каждая пара таких текстов используется однократно или небольшое число раз. Назовем такие методы анализа статистическими. Наиболее популярными статистическими методами являются дифференциальный, предложенный Бихамом и Шамиром [5] и линейный, впервые опубликованный Мацуи [6].

По отношению к статистическим методам можно организовать адаптивное противодействие. Снижающиеся оценки стойкости шифра ведут к снижению объема статистики, необходимого для вскрытия ключа. Поэтому в системе шифрованной связи нужно организовать плановую смену ключей и адаптивно (в соответствии с изменяющимися оценками стойкости) ограничивать объем

---

<sup>1</sup> Это наводит на мысль, что все эти способы криптоанализа в действительности были изобретены не за рубежом.

шифруемых данных на одном ключе, не позволяя нарушителю набрать требуемый объем статистики.

Обычно при анализе полагают, что для подстановочно-перестановочного шифра распределение вероятностей дифференциалов и линейных сумм подстановки на данном цикле не зависит от предыдущих циклов. Стойкость подстановочно-перестановочного шифра по отношению к линейному и дифференциальному методам растет в среднем экспоненциально от числа циклов. Действительно, на каждом цикле характеристика должна содержать хотя бы один активный  $S$ -блок, при этом вероятности дифференциалов (линейных сумм) активных  $S$ -блоков перемножаются, а влияние «параллельных» характеристик с общими активными входными и выходными разрядами существенно.

Например, если в шифре ГОСТ 28147–89 используются экстремальные подстановки (вероятности дифференциалов не превышают 0,25, абсолютные значения преобладаний для линейных сумм не превышают 0,25), то для вскрытия ключа обычным дифференциальным или линейным методом требуется недостижимый объем статистики [3].

Наряду со статистическими используются и алгебраические методы, не требующие большого объема статистики. Практически достаточно одного или нескольких открытых текстов и соответствующих шифртекстов. Предположим, что в блочном шифре отображение  $\{x\} \times \{k\} \rightarrow \{y\}$ , где  $x$ ,  $k$ ,  $y$  — соответственно открытый текст, ключ и шифртекст, является случайным. Используя аппроксимацию графа такого отображения критическим пуассоновским процессом, можно показать, что ключ определен однозначно, если объем открытых и соответствующих зашифрованных текстов превышает длину ключа в  $e/2 \approx 1,36$  раз [2].

Метрика между истинным и тестируемым ключом задается целевой функцией  $H$ . Предположим, что двоичные  $N$ -разрядные векторы открытого текста  $\mathbf{x}$  и шифртекста  $\mathbf{y}$  однозначно определяют  $n$ -разрядный ключ  $\mathbf{k}$ . Обозначим  $E_i(\mathbf{x}, \mathbf{k})$  — обратимый по первому аргументу оператор зашифрования на  $i$ -м цикле шифрования,  $r$  — число циклов шифрования. Определим промежуточные тексты

$$\mathbf{u} = \prod_{i=1}^{r/2} E_i(\mathbf{x}, \mathbf{k}), \quad \mathbf{v} = \prod_{i=r}^{r/2+1} E_i^{-1}(\mathbf{y}, \mathbf{k}),$$

полученные соответственно зашифрованием открытого текста на половине циклов и расшифрованием шифртекста на половине циклов. В силу обратимости оператора шифрования равенство  $\mathbf{u} = \mathbf{v}$

эквивалентно равенству  $\mathbf{y} = \prod_{i=1}^r E_i(\mathbf{x}, \mathbf{k})$  и достижимо тогда и только тогда, когда тестируемый ключ является истинным.

Один из первых алгебраических методов был предложен Д. Андельманом и Дж. Ридсом [4] для анализа дисковых шифраторов. Этот метод основывался на арифметическом продолжении булевых функций, заданных в базисе И, ИЛИ, НЕ, на множество вещественных чисел из интервала  $(0, 1)$ :  $a \& b \rightarrow ab$ ,  $a \vee b \rightarrow a + b$ ,  $\bar{a} \rightarrow 1 - a$ . Поскольку все булевы функции, описывающие шифрование, являются сбалансированными, то их продолжения на единичном  $n$ -мерном кубе  $(0, 1)^n$  принимают значения из интервала  $(0, 1)$ . Продолженная целевая функция принимает значение 1 (глобальный экстремум) тогда и только тогда, когда тестируемый ключ совпадает с истинным. Для поиска экстремума Андельман и Ридс предложили использовать аппарат дифференциального исчисления над полем вещественных чисел. Однако нарушение алгебраических свойств при переходе от булевых функций к арифметическим приводит к тому, что появляются многочисленные локальные экстремумы, число которых значительно превышает число ключей. Кроме того, сложность вычисления производной от многократной композиции функций растет по экспоненте от числа композиций. Практически для всех шифров сложность вычисления производной целевой функции превышает сложность перебора ключей.

Развитие метода Андельмана — Ридса [3] позволило устранить этот недостаток. Для вскрытия ключа можно определить арифметически продолженные булевы функции над множеством  $\{0; 0,5; 1\}$ . Если начальное приближение ключа близко к истинному значению, то поиск максимума целевой функции методом наискорейшего спуска позволяет вычислить оставшиеся биты ключа. Таким образом, стойкость каждого шифра обусловлена пороговым числом битов ключа, которые нужно угадать, а оставшиеся биты можно вычислить с полиномиальной сложностью.

Современные шифры разрабатываются так, чтобы обеспечивать стойкость и в том случае, если нарушитель знает открытые и соответствующие зашифрованные тексты (криптоанализ на основе известных открытых текстов), и даже может подбирать открытые тексты для зашифрования и по-

лучать соответствующие шифртексты (криптоанализ на основе подобранных открытых текстов).

## 2. Криптоанализ на основе продолженных многочленов Жегалкина

А. Г. Ростовцевым и Е. Б. Маховенко предложены новые методы криптоанализа, основанные на рациональном и 2-адическом продолжении многочленов Жегалкина [1].

Кольцо многочленов Жегалкина  $\mathbf{G}_n$  определяется следующим образом:

$$\mathbf{G}_n \cong \mathbf{F}_2[k_1, \dots, k_n]/(k_1^2 \oplus k_1, \dots, k_n^2 \oplus k_n).$$

В этом кольце каждая булева формула может быть единственным образом представлена многочленом Жегалкина.

Кольцо многочленов Жегалкина обладает следующими свойствами [3].

1. Для любого  $f \in \mathbf{G}_n$  справедливо  $f(f \oplus 1) = 0$ , то есть каждый непостоянный многочлен является делителем нуля. Поэтому каждый элемент кольца  $\mathbf{G}_n$  идемпотентен.
2. Единственным обратимым элементом кольца  $\mathbf{G}_n$  является 1. Поэтому кольцо частных  $\mathbf{G}_n$  совпадает с  $\mathbf{G}_n$ .
3. Каждый многочлен из  $\mathbf{G}_n$  обладает однозначным разложением на простые множители. Эквивалентными являются разложения  $f = 1 \cdot f$  и  $fg = f \cdot (fg)$ .
4. Неразложимыми элементами являются в точности те многочлены, которые на множестве из  $2^n$  наборов аргументов принимают значение 0 однократно. Поэтому существует  $2^n$  неразложимых многочленов, каждый из которых имеет степень  $n$ .
5. Каждый идеал в  $\mathbf{G}_n$  является главным.
6. Переход от булевой функции, заданной таблично, к многочлену Жегалкина, осуществляется умножением над  $\mathbf{F}_2$  вектора из  $2^n$  значений на матрицу Адамара.
7. Если вектору значений  $\mathbf{f}$  соответствует вектор коэффициентов  $\mathbf{a}$ , то вектору значений  $\mathbf{a}$  соответствует вектор коэффициентов  $\mathbf{f}$ .
8. Множество автоморфизмов кольца  $\mathbf{G}_n$  совпадает с множеством перестановок  $2^n$  неразложимых многочленов.

Продолжение кольца  $\mathbf{G}_n$  с неупорядоченного множества  $\mathbf{F}_2$  на упорядоченное множество  $\mathbf{Q}$  рациональных чисел с евклидовым нормированием и на множество  $\mathbf{Z}_2$  целых 2-адических чисел с 2-адическим неевклидовым нормированием позволяет определить вычислимую метрику, показывающую насколько далек тестируемый ключ от истинного.

Для вложения  $\mathbf{G}_n$  в соответствующую структуру над  $\mathbf{Q}$  используется продолжение [1]:

$$a \oplus b \rightarrow |a - b|, ab \pmod{2} \rightarrow ab. \quad (1)$$

Отметим, что в случае такого продолжения теряется ассоциативность в объемлющей структуре, но зато она имеет «характеристику» 2. Аргументы достаточно определить над множеством  $\{0, 0,5, 1\}$ .

2-адическое нормирование  $\text{val}$  целого числа определяется следующим образом:  $\text{val}(2^a b) = -a$ , где  $b$  — нечетное число. Для вложения  $\mathbf{G}_n$  в соответствующую структуру над  $\mathbf{Z}_2$  по аналогии с приближением вещественных чисел рациональными, когда отбрасываются младшие разряды, целое 2-адическое число можно приближенно представлять элементом кольца  $\mathbf{Z}/2^m \mathbf{Z}$ . В этом случае отбрасываются старшие двоичные разряды целого числа, минимальное значение нормирования равно  $-m$  и продолжение имеет вид

$$a \oplus b \rightarrow a + b \pmod{2^m}, ab \pmod{2} \rightarrow ab \pmod{2^m}. \quad (2)$$

Иногда сложение лучше продолжать так:

$$a \oplus b \rightarrow |a - b| \pmod{2^m}. \quad (2a)$$

Продолжение (2) обеспечивает гомоморфное вложение кольца  $\mathbf{G}_n$  в кольцо  $\mathbf{Z}/2^m \mathbf{Z}[k_1, \dots, k_n]/A_m$ , где  $A_m = \left( k_1^m (1 - k_1^{2^{m-2}}), \dots, k_n^m (1 - k_n^{2^{m-2}}) \right)$ . Однако в этом кольце нарушается ряд свойств  $\mathbf{G}_n$ , в частности, нет однозначности разложения на простые множители, идемпотентности и т. п. В случае продолжения (2a) объемлющая структура не является кольцом. Аргументы достаточно определить над множеством  $\{0, 1, 2\}$ . Продолжение (2) и (2a) приводит к тому, что одинаковые нули исходной целевой функции в  $\mathbf{G}_n$  становятся целыми 2-адическими числами, для которых определено отношение порядка. Применительно к целевой функции это продолжение можно уподобить «микроскопу», который позволяет различать значения продолженной целевой функции, соответствующие нулевому значению исходной целевой функции как многочлена Жегалкина.

В основе криптоанализа лежит процедура нахождения локального максимума продолженной

целевой функции, стартуя с начального приближения, в котором почти все разряды не определены (имеют значение 0,5 или 2). В основе метода лежит предположение, что истинное значение бита ключа обычно дает увеличение продолженной целевой функции. Этому соответствует эквивалентное утверждение, что ложному значению бита ключа обычно ведет к уменьшению целевой функции.

Пусть  $H^*$  — значение целевой функции для начального приближения. Разряды ключа, вычисленные в ходе поиска максимума целевой функции, такие что для значений 0 и 1 соответствующие значения целевой функции в одном случае меньше  $H^*$  а в другом — больше  $H^*$ , дают оценку соответствующих битов ключа.

Криптоанализ содержит три этапа. На первом этапе для произвольных ключей и начальных приближений, в которых почти все разряды ключа не определены. Для каждого разряда ключа в ходе поиска локального максимума целевой функции набирается статистика вида  $\{N_{ij}\}$ , где  $i, j \in \{0, 1\}$  и  $N_{ij}$  — число оценок данного бита ключа как  $i$ , если в действительности он равен  $j$ . Затем для каждого разряда ключа с помощью найденной статистики составляются матрицы условных вероятностей  $(q_{ij})$  того, что бит ключа в действительности равен  $i$ , если его оценка равна  $j$ . При этом  $q_{00} = \frac{N_{00}}{N_{00} + N_{01}}$ ,  $q_{01} = \frac{N_{10}}{N_{10} + N_{11}}$ ,  $q_{10} = \frac{N_{01}}{N_{00} + N_{01}}$ ,  $q_{11} = \frac{N_{11}}{N_{10} + N_{11}}$ . Пусть  $\mathbf{p} = (p_0, p_1)$  — вектор частот того, что оценка бита ключа равна  $i$ ,  $\boldsymbol{\pi} = (\pi_0, \pi_1)$  — вектор вероятностей того, что в действительности бит ключа равен  $i$ . Тогда имеет место равенство  $\boldsymbol{\pi} = (q_{ij})\mathbf{p}$ .

На втором этапе криптоанализа, предполагая, что матрицы  $(q_{ij})$  существенно не изменятся, если ключ неизвестен,<sup>2</sup> вычисляются векторы частот  $\mathbf{p}$  для каждого разряда ключа путем поиска максимума целевой функции. Затем вычисляются векторы  $\boldsymbol{\pi}$  и в соответствии с полученными вероятностями упорядочивается множество ключей по вероятности быть истинным. Значения (округления) вектора  $\boldsymbol{\pi}$  определяют *предположительное значение* разряда ключа.

На третьем этапе выполняется опробование ключей, начиная с наиболее вероятных. Обычно этот этап является наиболее трудоемким.

Если частоты  $\alpha = N_{00}/N_{10}$  и  $\beta = N_{01}/N_{11}$ , найденные на первом этапе анализа, различаются, то это обстоятельство тоже можно использовать для вскрытия ключа. Действительно, на втором этапе анализа можно найти частоту  $\gamma = N_{0^*}/N_{1^*}$ , где символом  $*$  обозначено неизвестное значение бита ключа. Нахождение предположительного значения бита ключа заключается в выборе одной из альтернативных гипотез:  $\gamma$  ближе к  $\alpha$ , чем к  $\beta$  или противоположной.

Неформально данный метод криптоанализа заключается в том, что на первом этапе для каждого бита ключа составляется статистический «нулевой и единичный портреты шифра» для некоторой выборки ключей и начальных приближений. На втором этапе для неизвестного ключа для выборки начальных приближений составляется аналогичный «портрет», и решается задача, на какой (нулевой или единичный) «портрет» он больше похож.

Ценностью данного метода является то, что он позволяет создать статистику оценок практически неограниченного объема на основании одного или нескольких открытых текстов. Для анализа статистики и ее связи с ключом могут использоваться и другие зависимости.

Предложенные методы криптоанализа были опробованы на подстановочно-перестановочном шифре с длиной блока и ключа по 64 бита. Каждый из 16 циклов шифрования содержал следующие операции.

1. Поразрядное сложение текста с ключом по модулю 2.
2. Перестановку битов в блоке  $x_i \rightarrow x_{23i \pmod{64}}$ .
3. Экстремальную подстановку  $S = (0, 13, 11, 8, 3, 6, 4, 1, 15, 2, 5, 14, 10, 12, 9, 7)$ .
4. Циклический сдвиг блока на  $25i$  разрядов влево, где  $i$  — номер цикла шифрования (в некоторых экспериментах использовался фиксированный сдвиг на 25 бит).

Вероятности дифференциалов подстановки не превышают 0,25, а все дифференциалы веса 2 имеют нулевую вероятность. Линейные суммы подстановки имеют абсолютные величины преобладаний не более 0,25, а линейные суммы веса 2 — не более 0,125. Никакие два разряда с выхода  $S$ -блока предыдущего цикла шифрования не попадают на вход одного и того же  $S$ -блока на следующем цикле шифрования. Поэтому стойкость данного шифра к линейному и дифференциальному методам анализа близка к переборной или даже превышает ее.

<sup>2</sup> Неформально это означает, что если оценка похожа на ключ при известном ключе, то она будет похожа на ключ и при неизвестном ключе.

Криптоанализ методом 2-адического продолжения для одиночного открытого текста показал, что при  $m = 128$  сложность вскрытия ключа практически не зависит от того, используется ли фиксированный или переменный циклический сдвиг, то есть от того, является ли шифр степенным. Усреднение по всем начальным приближениям и по всем разрядам ключа показало, что разряд ключа вскрывается правильно с частотой примерно 0,62. Метод рационального продолжения дал аналогичную статистику, но в целом для данного шифра он оказался несколько менее эффективным, чем метод 2-адического продолжения.

### 3. Стойкость ГОСТ 28147–89

Метод рационального продолжения был применен к отечественному стандарту шифрования ГОСТ 28147–89, который представляет собой 32-цикловый фейстелев шифр с длиной блока 64 бита и длиной ключа 256 бит. Этот шифр допускает надежное засекречивание информации с любым грифом секретности.

Для вскрытия ключа нарушитель может применять гипотетический вычислитель с производительностью  $10^{23}$  операций в год [3]. Если зашифрованная информация должна храниться в секрете 30 лет, а производительность вычислительной техники каждый год удваивается, то стойкость шифра должна составлять не менее  $10^{32}$  операций.

Шифрование данных по ГОСТ 28147–89 выполняется в режимах гаммирования или гаммирования с обратной связью, основанных на рассмотренном ниже режиме простой замены. На каждом цикле шифрования блок текста разбивается на левую и правую половины (слова) по 32 бита и выполняются следующие операции:

- сложение правого слова с ключом по модулю  $2^{32}$ ;
- 4-битовая подстановка для правого слова (вид блока подстановок не фиксирован);
- 11-битовый циклический сдвиг правого слова;
- фейстелева итерация.

Восемь 32-битовых слов ключа используются на 32 циклах шифрования следующим образом:  $K_1, \dots, K_8, K_1, \dots, K_8, K_1, \dots, K_8, K_8, \dots, K_1$ .

Пусть  $y = E(k, x)$  — уравнение шифрования ГОСТ 28147–89, где  $x, y, k$  соответственно открытый текст, шифртекст, ключ и  $\phi$  — инверсия старшего бита в каждом 32-разрядном слове указанных текстов и ключа. Тогда справедливо равенство  $\phi(y) = E(\phi(k), \phi(x))$  (это равенство можно интерпретировать как вычислимый автоморфизм шифра). Это свойство является аналогом свойства дополнения, присущего фейстелеву шифру, например, DES. Поэтому ключ можно искать перебором с точностью до старшего разряда.

К настоящему моменту неизвестны или не опубликованы методы криптоанализа ГОСТ 28147–89, снижающие его стойкость по сравнению с перебором по половине множества ключей. Можно показать, что наиболее популярные за рубежом линейный и дифференциальный методы криптоанализа применительно к отечественному шифру требуют недостижимого объема статистики. Криптоанализ на основе списка ключей применительно ГОСТ 28147–89 также не работает. Если использовать промежуточные двадцать циклов шифрования, так чтобы список ключей обладал центральной симметрией, то теоретически ключ поддается вскрытию, однако этот результат, очевидно, неприменим для 32-циклового шифра.

По отношению к анализу методом сдвига ГОСТ допускает слабые ключи, которые приводят к степенному уравнению шифрования с периодом повторения 1, 4 или 8 циклов шифрования. В общем случае для периода повторения из восьми циклов шифрования это условие можно записать в виде  $K_1 = K_8, K_2 = K_7, K_3 = K_6, K_4 = K_5$ . Вероятность слабого ключа пренебрежимо мала (примерно  $10^{-38}$ ). Кроме того, нахождение требуемой пары открытых текстов и соответствующих шифрограмм таких, что первый открытый текст после шифрования на восьми циклах равен второму открытому тексту (то же справедливо и для шифрограмм), по-видимому, не является легкой задачей. Поэтому метод сдвига тоже не работает. Поэтому можно считать, что стойкость ГОСТ 28147–89 равна  $2^{254} \approx 10^{76}$  вычислительных операций (или операций опробования).

Обычно для подстановочно-перестановочных шифров (к ним относится и ГОСТ 28147–89) рекомендуют выбирать подстановки, обеспечивающие максимальную стойкость по отношению к линейному и дифференциальному криптоанализу. Поэтому в эксперименте была выбрана подстановка  $S = (0, 13, 11, 8, 3, 6, 4, 1, 15, 2, 5, 14, 10, 12, 9, 7)$ , дифференциалы которой имеют вероятность не более 0,25, при этом дифференциалы веса 2 невозможны. Абсолютные значения преобла-

даний для линейных аппроксимаций входов и выходов подстановки не превышают 0,25, при этом линейные суммы веса 2 имеют абсолютные значения преобладаний не более 0,125. Это подстановка обеспечивает максимально возможную стойкость по отношению к дифференциальному и линейному криптоанализу.

Перемешивающие и рассеивающие свойства ГОСТ 28147–89 в значительной мере обусловлены переносами, возникающими при сложении текста с ключом. Влияние переносов уменьшается, если хотя бы одно из слагаемых является разреженным. Поэтому в целевой функции использовались разреженные открытые тексты (четыре блока, один из которых не содержал ни одной единицы, а остальные содержали по одной единице).

Программа криптоанализа написана на языке С, каждый разряд ключа и текстов, а также целевая функция представлялись типом “double” с плавающей точкой.

В ходе эксперимента исследовалась зависимость сложности криптоанализа от разреженности ключа: использовались ключи, содержащие по 8, 12 и 16 единиц в каждом 32-разрядном слове. Эксперимент содержал три этапа.

На первом этапе эксперимента вычислялись матрицы условных вероятностей ( $q_{ij}$ ) для  $>10^5$  случайных тестируемых ключей для начального приближения, в котором все разряды ключа равны 0,5. При этом число единиц в каждом 32-разрядном слове тестируемого ключа выбиралось равным соответственно 8, 12 или 16.

На втором этапе эксперимента выбирался случайный ключ соответствующей разреженности и для каждого разряда ключа на основе  $>10^5$  начальных приближений вычислялись векторы оценок  $\mathbf{p} = (p_0, p_1)$  и векторы  $\boldsymbol{\pi} = (\pi_0, \pi_1) = (q_{ij})\mathbf{p}$ . Затем проверялось число совпадений предположительных значений бита ключа как округления вектора  $\boldsymbol{\pi}$  с истинным значением (предположительное значение равно 0, если  $\pi_0 > 0,5$ ; предположительное значение равно 1, если  $\pi_1 > 0,5$ ).

Третий этап эксперимента выполнялся аналогично второму. Отличие заключалось в том, что использовалось 200 случайных ключей, содержащих по 16 единиц в слове ключа, и 512 начальных приближений для каждого ключа. Найденные векторы  $\boldsymbol{\pi} = (\pi_0, \pi_1)$  усреднялись для каждого разряда по всем ключам и начальным приближениям (суммарный объем статистики  $>10^5$ ).

В ходе всех этапов эксперимента наиболее часто вскрывались только биты первых трех слов ключа (96 бит), биты четвертого и последующих слов ключа вскрывались значительно реже. Поэтому полученные результаты справедливы для первых трех слов ключа.

По результатам первого и второго этапов эксперимента установлено, что если число единиц в каждом слове ключа равно 8 или 12, то почти для всех битов ключа выполнялось условие  $\pi_1 > \pi_0$  (это условие не выполнялось только для трех из 96 битов ключа). Достаточно большой объем статистики показывает, что, по-видимому, это условие можно считать надежным критерием того, что ключ является разреженным (по крайней мере, для используемой подстановки). Следовательно, свойство разреженности ключа является эффективно распознаваемым. Поэтому разреженные ключи можно считать слабыми.

Если число единиц в каждом слове ключа равно 16, то случаи  $\pi_1 > \pi_0$  и  $\pi_1 < \pi_0$  встречаются примерно одинаково часто. При этом округления для  $\pi_1, \pi_0$  дают правильные оценки для 57 из 96 битов ключа, что соответствует положительному преобладанию 0,09375. Это равносильно снижению шенноновской энтропии на один бит ключа с 1 до 0,974. Аналогично, значение  $\gamma$ , найденное на втором этапе, было ближе к правильному из двух значений  $\alpha$  и  $\beta$ , найденных на первом этапе, в 55 из 96 случаев, то есть имело положительное преобладание 0,0729.

Таким образом, для вскрытия ключа можно использовать оба критерия, предложенных во втором разделе: как по округлению вектора  $\boldsymbol{\pi}$ , так и по разности частот  $\alpha = N_{00}/N_{10}$  и  $\beta = N_{01}/N_{11}$ . Для обоих критериев преобладание положительное, это позволяет снизить сложность вскрытия ключа ГОСТ 28147–89 по сравнению с перебором.

Если найденные оценки для частот совпадения бита ключа с истинным значением будут справедливы при дальнейшем увеличении объема статистики, то для упрощения задачи вскрытия ключа можно использовать метод накопления.

Допустим, что справедливо предположение: вероятность правильного вскрытия бита ключа превышает 0,5 и не зависит от вида разреженных открытых текстов.

Объем эксперимента не позволяет утверждать, что это предположение справедливо, однако оно представляется правдоподобным. В случае справедливости предположения можно увеличить итоговое преобладание, используя несколько четверок разреженных открытых текстов.

Пусть для одиночного набора из четырех открытых текстов вероятность правильного определения бита ключа равна  $0,5 + \epsilon$ , где  $\epsilon > 0$ . В случае использования двух таких четверок (и соответственно двух наборов векторов  $\pi$ ) правильные биты ключа могут быть определены следующим образом. Бит ключа вероятно равен 0, если обе оценки для  $\pi_0$  превышают 0,5; бит ключа вероятно равен 1, если обе оценки для  $\pi_1$  превышают 0,5.

Предположим, что обе оценки для  $\pi_i = 0,5 + \epsilon > 0$  и совпадают. Вероятность того, что бит ключа определен неверно, равна вероятности двойной ошибки:  $p_{\text{ош}} = (0,5 - \epsilon)^2$ ; вероятность того, что бит ключа дважды определен правильно, равна вероятности:  $p_{\text{пр}} = (0,5 + \epsilon)^2$ . Эти два события исчерпывают всевозможные исходы. Итоговая вероятность правильного определения бита ключа равна

$$P_2 = \frac{(0,5 + \epsilon)^2}{(0,5 + \epsilon)^2 + (0,5 - \epsilon)^2}.$$

Аналогично, если используются  $m$  четверок разреженных текстов, то в случае совпадения  $m$  оценок вероятность того, что бит ключа будет вскрыт правильно, будет равна

$$P_m = \frac{(0,5 + \epsilon)^m}{(0,5 + \epsilon)^m + (0,5 - \epsilon)^m}.$$

Это позволяет получить гарантированно большую вероятность  $P_m$ . Для  $\epsilon = 0,05$  расчетная зависимость вероятности  $P_m$  и энтропии  $H$  на один бит ключа от объема статистики  $m$  представлена в таблице.

Вероятность правильного вскрытия бита ключа в зависимости от объема статистики.

$m$	1	2	3	5	10	20	30	50
$P$	0,55	0,599	0,771	0,732	0,881	0,982	0,9976	0,999956
$H$	0,993	0,971	0,938	0,839	0,381	0,1130	0,024	0,00070

Согласно таблице, при использовании нескольких десятков подобранных открытых текстов можно обеспечить гарантированно малую вероятность ошибочного определения бита ключа. В этом случае сложность криптоанализа определяется первыми двумя этапами и практически не превышает  $10^{10}$ , то есть является недопустимо низкой. По-видимому, увеличение вероятности  $P_m$  может быть получено также комбинацией методов рационального и 2-адического продолжения на меньшем объеме статистики. В общем случае оценка бита ключа может быть найдена как наиболее часто встречающееся значение округления для  $\pi_0$  ( $\pi_1$ ).

Следовательно, если указанное предположение верно, то оказывается справедливым по крайней мере одно из двух утверждений.

1. ГОСТ 28147–89 обладает слабыми подстановками, которые считаются наилучшими для подстановочно-перестановочных шифров и обеспечивают максимально возможную стойкость по отношению к линейному и дифференциальному криптоанализу.
2. Ключ ГОСТ 28147–89 может быть вскрыт с малой сложностью на основе нескольких десятков подобранных открытых текстов.

Таким образом, стойкость ГОСТ 28147–89 существенно зависит от того, выполняется ли указанное выше предположение.

### Библиографический список

1. Ростовцев А. Г., Маховенко Е. Б. Два подхода к анализу блочных шифров // Проблемы информационной безопасности. Компьютерные системы, СПб., 2002. № 1. С. 49–54.
2. Ростовцев А. Г., Маховенко Е. Б. Введение в криптографию с открытым ключом. — СПб.: Мир и Семья, Интерлайн, 2001.
3. Ростовцев А. Г., Маховенко Е. Б. Введение в теорию итерированных шифров. — СПб.: Мир и Семья, 2003.
4. Andelman D., Reeds J. On the cryptanalysis of rotor machines and substitution-permutation networks // IEEE transactions on information theory. 1982. V. IT–28. P. 578–584.
5. Biham E., Shamir A. Differential cryptanalysis of DES-like cryptosystems // Advances in Cryptology — CRYPTO '90. LNCS. Springer–Verlag. 1991. V. 537. P. 2–21.
6. Matsui M. Linear cryptanalysis method for DES cipher // Advances in Cryptology —

EUROCRYPT '93. 1994. LNCS. Springer-Verlag. V. 765. P. 386-397.