УДК 621.391:519.2

С.В. МИТИН

КОЛИЧЕСТВО ИНФОРМАЦИИ О КЛЮЧЕ, СОДЕРЖАЩЕЙСЯ В НАБОРАХ ОТКРЫТЫХ И ШИФРОВАННЫХ ТЕКСТОВ СИММЕТРИЧНОЙ РАНДОМИЗИРОВАННОЙ КРИПТОСИСТЕМЫ МАК-ЭЛИСА

Аннотация. Рассмотрена симметричная кодовая криптосистема, аналогичная рандомизированной (асимметричной) схеме шифрования Мак-Элиса. Получено выражение для количества информации о секретном ключе, которую можно извлечь из открытых и соответствующих им шифрованных сообщений криптосистемы. Показано, что при наличии этой информации стойкость симметричной криптосистемы к атакам на основе известного шифрованного текста совпадает со стойкостью ее асимметричного аналога.

Ключевые слова: криптография на основе кодов, схема шифрования Мак-Элиса, рандомизированная кодовая криптосистема, количество информации.

Одной из актуальных проблем современной криптографии является создание симметричных постквантовых криптосистем, стойкость которых базируется на сложности решения единственной вычислительной задачи подобно тому, как стойкость криптосистемы RSA базируется на сложности факторизации целых чисел. Перспективный класс для создания таких криптосистем образуют кодовые криптосистемы, и первой асимметричной из них является криптосистема Мак-Элиса [1].

В настоящее время известно несколько видов симметричных кодовых криптосистем, аналогичных криптосистеме Мак-Элиса. Это схемы шифрования Жордана [2], Рао [3] и криптосистема Рао–Нама [4], а также ряд ее усовершенствований [5, 6]. Однако ни одна из них полностью не удовлетворяет современным требованиям к стойкости и практичности одновременно.

Представленная статья посвящена исследованию симметричной кодовой криптосистемы, которая является аналогом рандомизированной схемы шифрования Мак-Элиса, предложенной в [7]. Определено, какую информацию о секретном ключе может извлечь противник, имея доступ к оракулу шифрования криптосистемы. Получено выражение для количества указанной информации. Показано, что при наличии этой информации стойкость симметричной криптосистемы к атакам на основе известного шифрованного текста совпадает со стойкостью ее асимметричного аналога.

Обозначим $\mathbf{N} = \{1, 2, \dots\}$ множество натуральных чисел, V_n — множество двоичных векторов длины n, $F_{m,n}$ — множество $m \times n$-матриц над полем $F = \mathbf{GF}(2)$, F_n^* — множество обратимых матриц порядка n над этим полем, $wt(x) = |\{i \in \overline{1,n} : x_i = 1\}|$ — вес вектора $x = (x_1, \dots, x_n) \in V_n$.

В дальнейшем термин (n, k, d)-код означает двоичный линейный код длины n, размерности k с минимальным расстоянием d.

Алгоритм декодирования (n, k, d)-кода C с параметром $t \le \lfloor 1/2 \cdot (d-1) \rfloor$ определяется как произвольный алгоритм, который получает на вход вектор $y = c \oplus e$ такой, что $c \in C$, $e \in V_n$, $wt(e) \le t$, и восстанавливает (единственную) указанную пару слов (c, e) по y.

Матрица $B \in F_{k\times n}$ ранга k называется специальной ступенчатой, если она имеет вид

$$B = \begin{pmatrix} & i_1 & i_2 & i_k \\ 0\ldots0 & 1*\ldots* & 0*\ldots* & 0*\ldots* \\ 0\ldots0 & 00\ldots0 & 1*\ldots* & 0*\ldots* \\ \ldots & \ldots & \ldots & \ldots \\ 0\ldots0 & 00\ldots0 & 00\ldots0 & 1*\ldots* \end{pmatrix},$$

где * обозначает произвольный элемент поля F, $1 \le i_1 < \ldots < i_k \le n$. Из известных теорем линейной алгебры (см., например, [8], разд. VII) следует, что каждый код C имеет единственную порождающую и единственную проверочную матрицы, которые являются специальными ступенчатыми. Указанные матрицы называются соответственно канонической порождающей и канонической проверочной матрицами кода C и их можно найти по произвольной системе векторов, порождающих код (как векторное пространство над полем F) с помощью алгоритма Гаусса.

Перейдем к определению симметричной рандомизированной криптосистемы с параметрами $l, k, n, t \in \mathbf{N}$, $p \in (0, 1/2)$, где $l \le k \le n$. Ключами этой криптосистемы являются упорядоченные наборы (S, M, P), где $S \in F_k^*$, M — порождающая матрица (n, k, d)-кода C с (быстрым) алгоритмом декодирования, P — подстановочная матрица порядка n.

Для шифрования открытого текста $m \in V_l$ на ключе (S, M, P) отправитель вычисляет матрицу $G = SMP$, генерирует независимые случайные векторы r и $\xi = (\xi_1, \ldots, \xi_n)$. В данном случае вектор r имеет равномерное распределение вероятностей на множестве V_{k-l}, а координаты вектора ξ являются независимыми случайными величинами, распределенными по закону $\mathbf{P}(\xi_i = 1) = 1 - \mathbf{P}(\xi_i = 0) = p$, $i \in \overline{1, n}$. Далее отправитель вычисляет шифрованный текст по формуле

$$E_G(m) = (m, r)G \oplus \xi. \tag{1}$$

Для расшифрования сообщения (1) на ключе (S, M, P) получатель вычисляет вектор $E_G(m)P^{-1}$ и применяет к нему алгоритм декодирования кода C с параметром t. Если $wt(\xi) \le t$, то получатель восстанавливает вектор $(m, r)S$, по которому, используя обратимость матрицы S, находит искомый открытый текст m.

В случае $wt(\xi) > t$ возможна ошибка расшифрования, вероятность возникновения которой можно сделать сколь угодно малой, выбрав надлежащим образом параметры криптосистемы. Отметим также, что единственное отличие между описанной криптосистемой и (асимметричной) рандомизированной криптосистемой Мак-Элиса [7] состоит в том, что матрица $G = SMP$, используемая в качестве открытого ключа критосистемы Мак-Элиса, секретная. При $l = k$ описанная криптосистема известна как схема шифрования Жордана–Рао [2, 3] и является менее стойкой (относительно атаки с подобранным открытым текстом), чем классическая криптосистема Мак-Элиса [1]. В [7] получены условия, при которых рандомизированная криптосистема Мак-Элиса (с открытым ключом) CPA-стойкая.

Предположим, что противник имеет доступ к оракулу (1), т.е. может передавать на вход функции E_G произвольные открытые тексты $m \in V_l$ и получать соответствующие шифрованные тексты $c = E_G(m)$. Выясним, какое количество информации о матрице G может получить противник. Для этого рассмотрим внача-

ле упрощенный вариант криптосистемы, в котором вектор ξ не используется для шифрования.

Итак, предположим, что матрица G, являющаяся секретным ключом, выбирается случайно и равновероятно из множества всех $k \times n$-матриц ранга k над полем F, а шифрование открытого текста $m \in V_l$ осуществляется по формуле

$$H_G(m) = (m, r)G, \tag{2}$$

где r — случайный равновероятный двоичный вектор длины $k-l$. Зная матрицу G, вследствие линейной независимости ее строк можно быстро восстановить векторы m и r по шифрованному тексту $H_G(m)$.

Отметим, что в случае $l = k = n$ описанная схема шифрования — это известный шифр Хилла [9], нестойкий относительно атак на основе подобранных открытых текстов, поскольку при наличии доступа к оракулу H_G можно восстановить матрицу G с помощью алгоритма Гаусса.

Рассмотрим задачу восстановления матрицы G для рандомизированного шифра, который определяется по формуле (2).

Утверждение 1. Пусть $G \in F_{k\times n}$, $\operatorname{rank}(G) = k$, G_1 и G_2 — подматрицы матрицы G, состоящие из ее первых l и последних $k-l$ строк соответственно, C — код, порожденный строками матрицы G_2. Тогда, имея доступ к оракулу (2), можно выполнить следующие операции:

— построить каноническую порождающую матрицу $G_{2,0}$ и каноническую проверочную матрицу $H_{2,0}$ кода C с вероятностью возникновения ошибки, не превышающей $\delta \in (0,1)$, использовав $N = k - l + \lceil \log \delta^{-1} \rceil$ запросов к оракулу и $O(\max\{N, n\}(\min\{N, n\})^2)$ двоичных операций;

— найти матрицу $G_1 H_{2,0}^{\mathrm{T}}$, использовав l запросов к оракулу и $O(nl(n-k+l))$ двоичных операций;

— найти матрицу $G_{1,0}$ такую, что

$$G_{1,0} H_{2,0}^{\mathrm{T}} = G_1 H_{2,0}^{\mathrm{T}}, \tag{3}$$

выполнив $O(nl(n-k+l)^2)$ двоичных операций.

Матрицу G можно представить в виде

$$G = \begin{pmatrix} I_l & X \\ 0 & U \end{pmatrix} \begin{pmatrix} G_{1,0} \\ G_{2,0} \end{pmatrix}, \tag{4}$$

где I_l — единичная матрица порядка l, $X \in F_{l\times(k-l)}$, $U \in F_{k-l}^*$. При этом любой набор сообщений $m_1, \ldots, m_t \in V_l$, $H_G(m_1), \ldots, H_G(m_t)$ не содержит дополнительной информации о G. Иными словами, при условии независимого случайного равновероятного выбора матриц X и U для любых $X_0 \in F_{l\times(k-l)}$, $U_0 \in F_{k-l}^*$, $m_1, \ldots, m_t \in V_l$ и $s_1 \in m_1 G_{1,0} \oplus C, \ldots, s_t \in m_t G_{1,0} \oplus C$ справедливо равенство

$$\mathbf{P}(X = X_0, U = U_0 \mid H_G(m_i) = s_i,\ i \in \overline{1, t}) =$$

$$= \mathbf{P}(X = X_0, U = U_0) = 2^{-l(k-l)} \left(\prod_{i=0}^{k-l-1} (2^{k-l} - 2^i) \right)^{-1}. \tag{5}$$

Доказательство. Из формулы (4) следует, что

$$H_G(m) = mG_1 \oplus rG_2,\ m \in V_l.$$

Итак, код C состоит из всех слов вида $H_G(0)$, а шифрование сообщения m заключается в случайном равновероятном выборе слова в смежном классе $mG_1 \oplus C$ кода C.

Для нахождения матриц $G_{2,0}$, $H_{2,0}$ передадим N раз на вход оракула (2) нулевое сообщение, в результате чего получим систему из N независимых случайных равновероятных слов кода C. Поскольку размерность этого кода равна $k-l$, вероятность того, что указанная система слов не порождает код C, равна отношению числа $N \times (k-l)$-матриц с линейно зависимыми столбцами над полем F к $2^{-N(k-l)}$, которое меньше, чем $2^{k-l-N} \le \delta$, в силу определения параметра N и леммы 2 в [10]. Итак, с вероятностью, не меньшей $1-\delta$, полученная система случайных слов кода C порождает этот код, и матрицы $G_{2,0}$, $H_{2,0}$ можно восстановить из этой системы с помощью алгоритма Гаусса, использовав $O(\max\{N, n\}(\min\{N, n\})^2)$ двоичных операций.

Заметим, поскольку строки каждой из матриц G_2 и $G_{2,0}$ образуют базис кода C, существует матрица $U \in F_{k-l}^*$ такая, что

$$G_2 = UG_{2,0}. \tag{7}$$

Подставляя (7) в (2) и используя равенство $G_{2,0}H_{2,0}^{\mathrm{T}} = 0$, получаем $H_G(m)H_{2,0}^{\mathrm{T}} = mG_1H_{2,0}^{\mathrm{T}}$, $m \in V_l$. Отсюда следует, что i-я строка матрицы $G_1H_{2,0}^{\mathrm{T}}$ равна $H_G(e_i)H_{2,0}^{\mathrm{T}}$, где e_i — двоичный вектор длины l, все координаты которого, за исключением i-й, равны нулю, $i \in \overline{1, l}$. Итак, матрицу $G_1H_{2,0}^{\mathrm{T}}$ можно вычислить за $O(nl(n-k+l))$ двоичных операций с помощью l запросов к оракулу (2).

Наконец, зная матрицы $H_{2,0}$ и $G_1H_{2,0}^{\mathrm{T}}$, находим матрицу $G_{1,0}$, удовлетворяющую равенству (3), за $O(nl(n-k+l)^2)$ двоичных операций, решая n систем линейных уравнений с одинаковой матрицей коэффициентов $H_{2,0}^{\mathrm{T}}$.

Поскольку $H_{2,0}$ является проверочной матрицей кода C, из формулы (3) следует, что существует матрица $X \in F_{l\times(k-l)}$ такая, что $G_1 = G_{1,0} \oplus XG_{2,0}$. Из полученного равенства и формулы (7) следует равенство (4).

Итак, для завершения доказательства остается убедиться в справедливости формулы (5).

Поскольку X и U являются независимыми случайными матрицами с равномерными распределениями на множествах $F_{l\times(k-l)}$ и F_{k-l}^* соответственно, для любых $X_0 \in F_{l\times(k-l)}$, $U_0 \in F_{k-l}^*$ имеем

$$\mathbf{P}(X = X_0, U = U_0) = |F_{l\times(k-l)}|^{-1}\,|F_{k-l}^*|^{-1} = 2^{-l(k-l)}\left(\prod_{i=0}^{k-l-1}(2^{k-l} - 2^i)\right)^{-1}. \tag{8}$$

Зафиксируем произвольные векторы $m_1, \ldots, m_t \in V_l$, $s_1 \in m_1G_{1,0} \oplus C, \ldots, s_t \in m_tG_{1,0} \oplus C$. Покажем, что при независимом случайном равновероятном выборе матриц X, U в формуле (4) и векторов $r_1, \ldots, r_t \in V_{k-l}$, используемых при вычислении сообщений $s_1, \ldots, s_t$ соответственно, справедливо равенство

$$\mathbf{P}(X = X_0, U = U_0 \mid H_G(m_i) = s_i, i \in \overline{1, t}) = \mathbf{P}(X = X_0, U = U_0). \tag{9}$$

Действительно, существуют векторы $c_1, \ldots, c_t \in C$ такие, что $s_i \in m_iG_{1,0} \oplus c_i$, $i \in \overline{1, t}$. Итак, на основании формул (2), (4), линейной независимости строк матри-

цы $G_{2,0}$ и обратимости случайной матрицы U справедливы равенства

$$\mathbf{P}(H_G(m_i) = s_i, i \in \overline{1, t}) = \mathbf{P}((m_i X \oplus r_i U)G_{2,0} = c_i, i \in \overline{1, t}) = 2^{-t(k-l)}.$$

Аналогично получим, что

$$\mathbf{P}(X = X_0, U = U_0, H_G(m_i) = s_i, i \in \overline{1, t}) = 2^{-t(k-l)}\,\mathbf{P}(X = X_0, U = U_0).$$

Итак, справедлива формула (9).

Таким образом, на основании равенств (8), (9) справедлива формула (5). Утверждение 1 доказано.

Назовем $k \times n$-матрицы G и G' ранга k эквивалентными, если существуют матрицы $X \in F_{l\times(k-l)}$ и $U \in F^*_{k-l}$ такие, что $G = \begin{pmatrix} I_l & X \\ 0 & U \end{pmatrix} G'$.

Полученное утверждение 1 показывает, что, имея доступ к оракулу (2), можно восстановить ключ G рандомизированного шифра Хилла лишь с точностью до класса эквивалентности, которому он принадлежит, т.е. с точностью до матриц X и U в формуле (4).

Таким образом, условная энтропия ключа при наличии любого набора открытых и соответствующих им шифрованных сообщений составляет

$$\log\left(2^{l(k-l)}\left(\prod_{i=0}^{k-l-1}(2^{k-l} - 2^i)\right)\right) \approx k(k-l) \text{ бит},$$

при этом безусловная энтропия ключа $\log\left(\prod_{i=0}^{k-1}(2^n - 2^i)\right) \approx kn$ битам.

Следствие. Имея доступ к оракулу (1), можно получить ровно $\log\left(2^{l(k-l)}\left(\prod_{i=0}^{k-l-1}(2^{k-l} - 2^i)\right)\right) \approx k(k-l)$ бит информации о ключе G симметричной рандомизированной криптосистемы Мак-Элиса, а именно указать класс эквивалентности, которому принадлежит этот ключ.

Отметим, что сообщение (1) является результатом искажения сообщения (2), причем вектор искажений $\xi = (\xi_1, \ldots, \xi_n)$ не зависит от последнего сообщения. Отсюда следует, что количество информации о ключе G, содержащейся в произвольном наборе открытых текстов $m_1, \ldots, m_t$ и соответствующем наборе шифрованных текстов $E_G(m_1), \ldots, E_G(m_t)$, не превышает количества информации об этом ключе, содержащейся в наборе $m_1, \ldots, m_t$, $H_G(m_1), \ldots, H_G(m_t)$. Для завершения доказательства остается применить полученное утверждение 1.

В заключение отметим, что, зная класс эквивалентности, которому принадлежит ключ G симметричной рандомизированной криптосистемы Мак-Элиса, можно построить на нее атаку на основе известного шифрованного текста, исходя из аналогичной атаки на соответствующую асимметричную криптосистему.

Действительно, пусть $y = (m, r)G \oplus \xi$ является шифрованным текстом, полученным из открытого текста m на секретном ключе $G = SMP$, где $S \in F^*_k$, M — образующая матрица некоторого (n, k, d)-кода C, а P — подстановочная матрица порядка n. Если известен класс эквивалентности, которому принадлежит ключ G, т.е. матрица $G_0 = \begin{pmatrix} G_{1,0} \\ G_{2,0} \end{pmatrix}$ в формуле (4), то на основании этой формулы справедливо равенство $y = (m, mX \oplus rU)G_0 \oplus \xi$, из которого следует, что y является результатом шифрования открытого текста m на открытом ключе $G_0 = \left(\begin{pmatrix} I_l & X \\ 0 & U \end{pmatrix}^{-1} S\right) MP$ асимметрич-

ной рандомизированной криптосистемы Мак-Элиса. Таким образом, любую атаку на последнюю криптосистему, позволяющую восстанавливать m по y, можно применить для нахождения открытого текста по шифрованному тексту в симметричной криптосистеме.

СПИСОК ЛИТЕРАТУРЫ

1. McEliece R.J. A public-key cryptosystem based on algebraic coding theory. *Prog. Rep., Jet Prop. Lab.,* California Inst. Technol., 1978. P. 114–116.
2. Jordan J.P. A variant of public-key cryptosystem based on Goppa codes. *Sigact news*. 1983. Vol. 15, N 1. P. 61–66.
3. Rao T.R.N. Cryptosystems using algebraic codes. In: *Proc. Int. Conf on Computer Systems & Signal Processing.* Dec. 84. Bangalore, India, 1984.
4. Rao T.R.N., Nam K.H. Private-key algebraic code encryption. *IEEE Trans. on Inform Theory*. 1987. IT-35(4). P. 829–833.
5. Sobhi Afshar A.A., Eghlidos T., Aref M.R. Efficient secure channel coding based on quasi-ciclic low-density parity-check codes. *Journal of IET-Communications*. 2009. Vol. 3, Iss. 2. P. 279–292.
6. Hooshmand R., Eghlidos T., Aref M.R. Improving the Rao-Nam secret key cryptosystem using regular EDF-QC-LDPC codes. *ISC Journal of Information sequrity*. 2012. Vol. 4, Issue 1. P. 3–14.
7. Nojima R., Imai H., Kobara K., Morozov K. Semantic security for the McEliece cryptosystem without random oracles. *Des. Codes Cryptography*. 2008. Vol. 49, N 1–3. P. 289–305.
8. Глухов М.М., Елизаров В.П., Нечаев А.А. Алгебра. Учебник в 2-х т., Т. 1. Москва: Гелиос АРВ, 2003. 336 с.
9. Stinson D.R. Cryptography: Theory and practice. Third Edition. CRC Press. 2005. 616 p.
10. Алексейчук А.Н., Конюшок С.Н. О статистических свойствах нелинейности сужений булевых функций на случайно выбранное подпространство. *Прикладная дискретная математика*. 2012. Вып. 1(15). С. 5–10.

Надійшла до редакції 18.11.2019

С.В. Мітін

КІЛЬКІСТЬ ІНФОРМАЦІЇ ПРО КЛЮЧІ, ЩО МІСТИТЬСЯ В НАБОРАХ ВІДКРИТИХ І ШИФРОВАНИХ ТЕКСТІВ СИМЕТРИЧНОЇ РАНДОМІЗОВАНОЇ КРИПТОСИСТЕМИ МАК-ЕЛІСА

Анотація. Розглянуто симетричну кодову криптосистему, аналогічну рандомізованій (асиметричній) схемі шифрування Мак-Еліса. Отримано вираз для кількості інформації про секретний ключ, яку можна видобути з відкритих і відповідних їм шифрованих повідомлень криптосистеми. Показано, що за наявності цієї інформації стійкість симетричної криптосистеми до атак на основі відомого шифрованого тексту збігається зі стійкістю її асиметричного аналога.

Ключові слова: криптографія на основі кодів, схема шифрування Мак-Еліса, рандомізована кодова криптосистема, кількість інформації.

S.V. Mitin

AMOUNT OF KEY INFORMATION CONTAINED IN OPEN AND ENCRYPTED TEXT SETS OF THE SYMMETRIC RANDOMIZED McELIECE CRYPTOSYSTEM

Abstract. A symmetric code cryptosystem, which is similar to the randomized (asymmetric) McEliece encryption scheme, is considered. An expression for the amount of information about the secret key, which can be extracted from the open and the corresponding encrypted messages of the cryptosystem, is obtained. It is shown that with this information, the security of the symmetric cryptosystem to the attacks based on known ciphertext coincides with the security of its asymmetric counterpart.

Keywords: code-based cryptography, McEliece encryption scheme, randomized code cryptosystem, amount of information.

Митин Сергей Вячеславович,
старший преподаватель Института специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского», e-mail: meetser@gmail.com.