УДК 621.383

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПЕРСОНАЛЬНЫХ ДАННЫХ

А. М. ТИМОФЕЕВ, *доцент кафедры защиты информации, к. т. н., доцент БГУИР*

В работе предложена криптосистема для зашифрования персональных данных и расшифрования шифротекстов. Описаны принципы функционирования данной криптосистемы, а также обосновано количество двоичных символов в одном шифруемом блоке, предложены методики оценки числа итераций криптосистемы и эффективности шифрования персональных данных.

ВВЕДЕНИЕ

В современном информационном обществе возникает острая потребность формирования, хранения, обработки, передачи и приема различной конфиденциальной информации [1–3], в частности, персональных данных.

Персональные данные – это любая информация, относящаяся к идентифицированному физическому лицу или физическому лицу, которое может быть идентифицировано [4].

На всех этапах жизненного цикла персональных данных крайне важно обеспечивать их информационную безопасность, в частности, скрытность от незаконного ознакомления и использования, что регулируется в Республике Беларусь соответствующим законодательством [4–12]. Для этого могут быть использованы методы введения идентификаторов, изменения состава, декомпозиции и перестановки [12]. Важно отметить, что криптографические методы и средства зашифрования персональных данных, позволяющие выполнить их обезличивание, входят в состав обязательных по обеспечению защиты персональных данных и являются одними из наиболее эффективных по сравнению с другими [1–3, 12].

При реализации криптографических методов зашифрования персональных данных целесообразно использовать симметричные одноключевые криптосистемы (с секретным ключом), которые характеризуются более высокой скоростью зашифрования и расшифрования в сравнении с асимметричными двухключевыми криптосистемами (с открытым ключом) [1–3]. В основном это обусловлено тем, что криптографические операции, используемые в симметричных одноключевых криптосистемах, являются наиболее простыми для современных аппаратных, программных и аппаратно-программных комплексов [1–3]. Однако использование симметричных одноключевых криптосистем для обеспечения конфиденциальности персональных данных может приводить к появлению избыточности зашифрованных данных. Это имеет место в случае блочного шифрования данных в стандартных схемах шифрования данных и расшифрования шифротекстов (стандарты DES, AES, ГОСТ 28147-89 и др. [1–3]), когда данные разбивают на блоки заданного размера. Если шифруемые данные, подлежащие зашифрованию, не кратны по длине размеру стандартного блока, то недостающие биты дополняются нулями до полного размера блока. Например, в стандарте AES размер одного блока равен 128 бит [13]. Если длина данных в одном шифруемом блоке открытого текста составляет 100 бит, то необходимо добавить 28 недостающих бит нулевыми значениями. В этом случае формируется блок данных длиной 128 бит, который затем зашифровывается. В результате получают шифротекст длиной 128 бит. Значащими битами при расшифровании такого шифротекста по-прежнему являются 100 бит, а оставшиеся 28 бит (нулевые значения) будут избыточными, которые не содержат полезной информации. Таким образом, снижается эффективность функционирования стандартной симметричной одноключевой криптосистемы, а также уменьшается эффективность использования пропускной способности канала связи при передаче и приеме шифротекстов. Разработка нового метода (алгоритма, стандарта) шифрования блока заданной длины (для описанного выше примера – блока размером 100 бит) может быть нецелесообразной, поскольку потребует значительных трудозатрат и становится непригодна при небольшом видоизменении размера блока. Например, персональные данные различных физических лиц могут отличаться по своей длине, в результате чего размер данных, подлежащих зашифрованию, тоже может измениться либо в меньшую, либо в бóльшую сторону (для описанного выше примера соответственно – стать либо меньше 100 бит, либо больше 100 бит).

В связи с тем, что применение стандартных схем шифрования данных и расшифрования шифротекстов блочного типа для обеспечения конфиденциальности персональных данных, свободных от указанных выше недостатков, развито не в полной мере, это являлось целью данной работы.

В качестве объекта исследования использовался криптографический алгоритм AES-128 [13]. Данный алгоритм выбран в связи с тем, что имеет достаточную криптостойкость для решения различных практических задач в области информационной безопасности, характеризуется высокими скоростями расширения ключа и шифрования/расшифрования при программной реализации, а также может быть реализован на базе смарт-карт как с большим, так и с ограниченным объемом ресурсов [1–3, 13].

Предметом исследований являлось применение криптоподобных преобразований информации для обеспечения конфиденциальности персональных данных, при которых эффективность функционирования канала передачи и приема зашифрованных персональных данных максимальна.

СТРУКТУРНАЯ СХЕМА БЛОКА ШИФРОВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ

На рисунке приведена структурная схема блока шифрования персональных данных. Рассмотрим принцип функционирования данной схемы. Предположим, что защищаемая информация представлены в виде базы данных, содержащей следующие сведения: дату и время ее создания, а также непосредственно персональные данные физического лица, подлежащие заполнению (см. рис. 1).

После формирования всей защищаемой информации вначале ее кодируют. Для этого воспользуемся системой кодировки UTF-16 [14, 15], которая используется в настоящее время в современных версиях операционных систем (например, Microsoft Windows) и не накладывает ограничений на прикладные программы касательно кодирования текстовых файлов, позволяя им использовать как UTF-16LE, так и UTF-16BE посредством установки и трактовки соответствующей метки порядка байтов. Отметим, что в файловых системах NTFS, а также FAT с поддержкой длинных имен, имена файлов записываются также

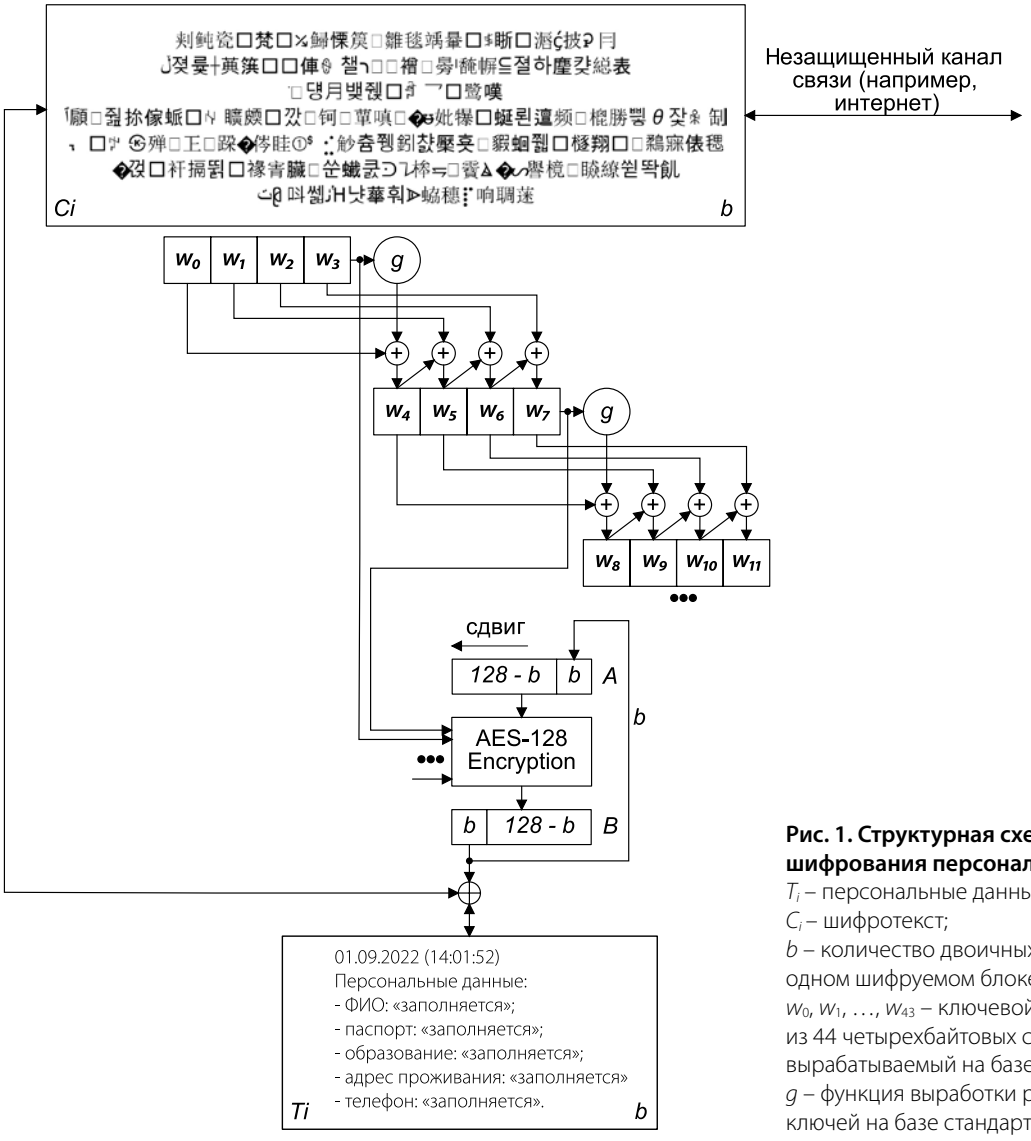

Рис. 1. Структурная схема блока шифрования персональных данных
T_i – персональные данные;
C_i – шифротекст;
b – количество двоичных символов в одном шифруемом блоке;
$w_0, w_1, \ldots, w_{43}$ – ключевой массив из 44 четырехбайтовых слов, вырабатываемый на базе стандарта [13],
g – функция выработки раундовых ключей на базе стандарта [13]

в UTF-16LE [14, 15]. В качестве примера в таблице 1 приведены результаты расчетов для первой строки персональных данных, подлежащих зашифрованию.

После этого в блоки $w_0 \div w_3$ загружается криптографический ключ CipherKey, а в регистр A записывается вектор инициализации IV [13]. Отметим, что генерация криптографического ключа может быть выполнена на базе любого алгоритма, порождающего последовательность чисел заданного размера (для рассматриваемого примера 128 бит), элементы которой почти независимы друг от друга и подчиняются заданному распределению (обычно дискретному равномерному) [1-3]. На основе CipherKey формуют набор раундовых криптографических ключей в соответствии с [13]. Результаты расчетов раундовых ключей приведены в табл. 2.

Пусть количество двоичных символов в одном шифруемом блоке $b = 16$ бит. Отметим, что число b можно и выбирать и другим, отличным от 16 бит. Однако при этом длину T_i целесообразно выбирать кратным b; в противном случае эффективность использования пропускной способности канала связи при передаче и приеме шифротекстов уменьшится.

Таблица 2. Результаты расчетов раундовых ключей [13] для схемы, показанной на рисунке

Roundkeys #	Результаты расчетов
1	9FEF32B55E127BD2530E95A48CADCD9E
2	085239D156404203054ED7A789E31A39
3	1DF02B764BB069754EFEBED2C71DA4EB
4	B1B9C2B0FA09ABC5B4F7151773EAB1FC
5	2671723FDC78D9FA688FCCED1B657D11
6	4B8EF09097F6296AFF79E587E41C9896
7	97C860F9003E4993FF47AC141B5B3482
8	2ED073562EEE3AC5D1A996D1CAF2A253
9	BCEA9E229204A4E743AD3236895F9065
10	458AD385D78E7762942345541D7CD531

Примечание: значения представлены в шестнадцатеричном формате при CipherKey = 9485B22BC1FD49670D1CEE76DFA3583A

Таблица 1. Результаты расчетов для строки #1 персональных данных, представленных на рис. 1 [14, 15]

Персональные данные			Полное наименование символа
Символ	HEX формат	Binary формат	
0	0030	0000000000110000	DIGIT ZERO
1	0031	0000000000110001	DIGIT ONE
.	002E	0000000000101110	FULL STOP
0	0030	0000000000110000	DIGIT ZERO
9	0039	0000000000111001	DIGIT NINE
.	002E	0000000000101110	FULL STOP
2	0032	0000000000110010	DIGIT TWO
0	0030	0000000000110000	DIGIT ZERO
2	0032	0000000000110010	DIGIT TWO
2	0032	0000000000110010	DIGIT TWO
	0020	0000000000100000	SPACE
(	0028	0000000000101000	LEFT PARENTHESIS
1	0031	0000000000110001	DIGIT ONE
4	0034	0000000000110100	DIGIT FOUR
:	003A	0000000000111010	COLON
0	0030	0000000000110000	DIGIT ZERO
1	0031	0000000000110001	DIGIT ONE
:	003A	0000000000111010	COLON
5	0035	0000000000110101	DIGIT FIVE
2	0032	0000000000110010	DIGIT TWO
)	0029	0000000000101001	RIGHT PARENTHESIS
	000A	0000000000001010	<control>

Примечание: HEX и binary – шестнадцатеричный и двоичный формат соответственно.

Число итераций схемы равно:

$$i_N = T_N/b, \quad (1)$$

где T_N – количество двоичных бит персональных данных.

Для примера, показанного на рис. 1, $T_N = 2896$ бит, поэтому

$$i_N = 2896/16 = 181. \quad (2)$$

Таким образом, с учетом используемой системы кодировки в рассматриваемом примере число итераций схемы i_N будет соответствовать числу символов, содержащихся в персональных данных, подлежащих заполнению. Следовательно, за каждую итерацию формируется 16 бит шифротекста.

Далее на основании данных из кодировочных таблиц формируются блоки T_i по 16 бит каждый (i – номер итерации). Затем запускается процесс шифрования персональных данных на базе режима CFB [1-3, 16], который реализуется следующим образом.

Для первой итерации ($i = 1$) вектор инициализации IV, содержащийся в регистре A, зашифровывается блоком AES-128 Encryption [13] (см. рис. 1). Полученное в результате значение записывается в регистр B и используется для вычисления первого блока шифротекста C_1. Остальные итерации ($i = 2 \div 181$) выполняются аналогично за исключением следующего. Вначале каждой итерации из регистра A удаляют b старших двоичных символов, оставшиеся биты сдвигают влево на b двоичных символов. В высвободившиеся младшие разряды регистра A записывают b двоичных бит шифротекста C_{i-1}, полученных на предыдущей итерации.

Шифротекст рассчитываем по формуле

$$C_i = T_i \oplus R_{B_i}, \quad (3)$$

где T_i – i-й блок шифруемого сообщения; R_{B_i} – b старших двоичных символов, содержащихся в регистре B после i-й итерации.

Так, например, при IV = 5349670D1CEE76DFA3583A9485B22BC1 для первой итерации

$$C_i = T_i \oplus R_{B_i} = 0000000000110000_2 \oplus 0101001000001110_2 = 0101001000111110_2 = 523E_{16}$$

Теперь закодируем полученный символ шифротекста, воспользовавшись системой кодировки UTF-16 [14, 15]. В результате первый символ шифротекста $C_1 = 523E_{16}$ в символьном представлении запишется как 刾. Аналогично зашифровываются и кодируются остальные персональные данные, показанные на рисунке.

Криптосистема, реализующая алгоритм расшифрования, имеет тот же вид, что и при зашифровании (см. рис.). При этом персональные данные расшифровываются по формуле [1–3, 16]

$$T_i = C_i \oplus R_{B_i}. \quad (4)$$

ОЦЕНКА ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО КАНАЛА СВЯЗИ

Для оценки эффективности функционирования разработанного криптографического канала связи будем использовать величину

$$E_C = \frac{T_N}{C_N} \times 100\ \%, \quad (5)$$

где C_N – количество двоичных бит шифротекста.

При зашифровании персональных данных, представленных на рис. 1, $E_C = (2896/2896) \times 100 = 100\ \%$. Следовательно, пропускная способность канала связи, используемого для передачи шифротекста, будет использоваться в полной мере. Для сравнения, при зашифровании тех же персональных данных, но с помощью стандартных блочных алгоритмов, например, DES и AES-128 соответственно в режимах ECB [1-3, 13, 16] величина E_C не достигает максимального значения и равна $E_C = (2896/2944) \times 100 \approx 98{,}37\ \%$. Это означает, что шифротекст, передаваемый по каналу связи, будет содержать только ≈ 98,37 % полезной информации (персональные данные), а оставшиеся ≈ 1,63 % – избыточная информация, которая никаких данных не содержит (будет иметь подряд следующие двоичные символы «0»), однако потребует соответствующей пропускной способности канала связи, снизив эффективность его использования. Отметим, что в случае более коротких персональных данных описанный выше эффект будет проявляться еще в большей мере.

Результаты, полученные в настоящей работе, могут быть использованы при реализации обезличивания персональных данных.

ЗАКЛЮЧЕНИЕ

На основании полученных результатов можно сделать следующие выводы:

1. Предложена структурная схема блока зашифрования персональных данных. Эта схема выполнена на базе стандартного блочного алгоритма

AES-128, который может быть заменен на любой другой блочный алгоритм в зависимости от предъявляемых криптографических требований. При этом общие принципы функционирования криптосистемы останутся прежними.

2. Предложенная криптосистема может быть реализована программно, аппаратно или программно-аппаратно, поскольку выполняется на основе

достаточно простых математических операций (сложение по модулю два, циклические сдвиги и пр.). Это дает возможность организовывать защищенные от несанкционированного доступа каналы связи в режиме реального времени, в том числе, используя аппаратные платформы от 8-битных до 64-битных.

3. Реализовав один пакет программного обеспечения, выполняющий зашифрование персональных данных, можно использовать этот же пакет для расшифрования шифротекстов, что упрощает криптосистему, в сравнении с асимметричными криптосистемами и сохраняет преимущества, свойственные стандартным симметричным одноключевым криптосистемам.

ЛИТЕРАТУРА

1. Милославская, Н. Г. Научные основы построения центров управления сетевой безопасностью в информационно-телекоммуникационных сетях / Н. Г. Милославская. – Москва: Горячая линия-Телеком, 2021. – 432 с.
2. Бутакова, Н. Г. Криптографические методы и средства защиты информации: учебное пособие / Н. Г. Бутакова, Н. В. Федоров. – СПб.: Интермедия, 2020. – 380 с.
3. Boyd, C. Protocols for Authentication and Key Establishment. – 2nd Edition / C. Boyd, A. Mathuria, D. Stebila. – Springer: Berlin, 2020. – 519 p.
4. О защите персональных данных [Электронный ресурс]: Закон Респ. Беларусь от 7 мая 2021 г. № 99-З. – Режим доступа: https://pravo.by/document/?guid=12551&p0=H12100099&p1=1. – Дата доступа: 11.02.2024.
5. О мерах по совершенствованию защиты персональных данных [Электронный ресурс]: Указ Президента Респ. Беларусь от 28 октября 2021 г. № 422. – Режим доступа: https://pravo.by/document/?guid=12551&p0=P32100422&p1=1. – Дата доступа: 11.02.2024.
6. О государственном информационном ресурсе «Реестр операторов персональных данных» [Электронный ресурс]: приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 1 июня 2022 г. № 94. – Режим доступа: https://pravo.by/document/?guid=12551&p0=T62205021&p1=1. – Дата доступа: 11.02.2024.
7. Об обучении по вопросам защиты персональных данных [Электронный ресурс]: приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 12 ноября 2021 г. № 194. – Режим доступа: https://pravo.by/document/?guid=12551&p0=T62104892&p1=1. – Дата доступа: 11.02.2024.
8. Об использовании Национальным центром защиты персональных данных средств от приносящей доходы деятельности и из иных источников, не запрещенных законодательством [Электронный ресурс]: приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 27 сентября 2022 г. № 140. – Режим доступа: https://pravo.by/document/?guid=12551&p0=T62205152&p1=1. – Дата доступа: 11.02.2024.
9. Об утверждении Инструкции о порядке обезличивания персональных данных лиц, которым оказывается медицинская помощь [Электронный ресурс]: постановление Министерства здравоохранения Республики Беларусь от 28 мая 2021 г. № 64 – Режим доступа: https://pravo.by/document/?guid=12551&p0=W22136916&p1=1. – Дата доступа: 11.02.2024.
10. Об утверждении Инструкции о порядке предоставления информации о персональных данных пассажиров органам пограничной службы и иным государственным органам, осуществляющим оперативно-розыскную деятельность [Электронный ресурс]: постановление Министерства транспорта и коммуникаций Республики Беларусь от 20 июня 2016 г. № 27. – Режим доступа: https://pravo.by/document/?guid=12551&p0=W21631085&p1=1. – Дата доступа: 11.02.2024.
11. О формах и порядке дачи и отзыва согласия на внесение и обработку персональных данных пациента [Электронный ресурс]: постановление Министерства здравоохранения РеспубликиБеларусь от 7 июня 2021 г. № 74. – Режим доступа: https://pravo.by/document/?guid=12551&p0=W22136940&p1=1. – Дата доступа: 11.02.2024.
12. О технической и криптографической защите персональных данных» [Электронный ресурс]: приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 12 ноября 2021 г. № 195. – Режим доступа: https://pravo.by/document/?guid=12551&p0=T62104893&p1=1. – Дата доступа: 11.02.2024.
13. Federal Information Processing Standards Publication 197. Advanced Encryption Standard (AES) [Электронный ресурс]. – Режим доступа: https://nvlpubs.nist.gov/nistpubs/fips/nist.fips.197.pdf. – Дата доступа: 27.11.2022.
14. ISO/IEC 10646:2020. Information Technology – Universal Coded Character Set (UCS) [Электронный ресурс]. – Режим доступа: https://www.iso.org/ru/standard/76835.html. – Дата доступа: 27.11.2022.
15. IETF RFC 2781. UTF-16, an Encoding of ISO 10646 [Электронный ресурс]. – Режим доступа: https://www.ietf.org/rfc/rfc2781.txt. – Дата доступа: 27.11.2022.
16. Federal Information Processing Standards Publication 46-3. Data Encryption Standard (DES) [Электронный ресурс]. – Режим доступа: https://csrc.nist.gov/csrc/media/publications/fips/46/3/archive/1999-10-25/documents/fips46-3.pdf. – Дата доступа: 27.11.2022.
17. Национальный центр защиты персональных данных Республики Беларусь. Правовая основа. Методологические документы. Пункт «10. Осуществление технической и криптографической защиты персональных данных в соответствии с классификацией информационных ресурсов (систем), содержащих персональные данные» [Электронный ресурс]. – Режим доступа: https://cpd.by/pravovaya-osnova/metodologicheskiye-dokumenty-rekomendatsii/. – Дата доступа: 11.02.2024.

A cryptosystem for encrypting personal data and decrypting ciphertexts is proposed in the work. The operating principles of this cryptosystem have been described. The number of binary symbols in one encrypted symbol, methods for estimating the number of iterations of the cryptosystem and the effectiveness of encryption of personal data were justified.

Получено 11.02.2024.