e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2020. № 1
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2020. № 1

UDC 004.058.5

METHOD OF DATA DEPERSONALIZATION IN PROTECTED AUTOMATED INFORMATION SYSTEMS

Spevakov A. G. – PhD, Associate Professor of the Information Security Department, Southwest State University, Kursk, Russian Federation.

Spevakova S. V. – Post-graduate student of the Computer Science Department, Southwest State University, Kursk, Russian Federation.

Primenko D. V. – Post-graduate student of the Computer Science Department, Southwest State University, Kursk, Russia.

ABSTRACT

Context. The problem of data depersonalization in information systems is considered. The analysis of modern approaches to depersonalization of data is carried out, it is revealed and proved by need of creation of the new method allowing to increase security of the processed data and their reliability. The object of the study was a model of data depersonalization, allowing to reduce the cost of protecting information systems.

Objective. The goal of the work is the analysis of modern methods of depersonalization and the creation of a method that eliminates the identified shortcomings, with an increased level of confidentiality and use of hashing of critical data and a private key.

Method. A method of personal data depersonalization is proposed, based on the method of entering identifiers using hashing of critical data and a private key, which allows to increase the confidentiality of information processed in information systems. Methods are proposed for selecting key critical attributes from primary documents that uniquely identify the subject of personal data, the method of generating initial sets, which divides the source data into two disjoint subsets, the method of generating a hash identifier from a unique sequence and a private key that depersonalizes information and enhances its confidentiality.

Results. The developed method is implemented in software and researched while solving the problems of depersonalization.

Conclusions. The carried out experiments confirmed the efficiency of the proposed method and allow to recommend it for implementation in automated information systems for processing personal data for solving problems of depersonalization. Prospects for further research may be in the creation of hardware streamlined data depersonalization allowing to increase the speed of processing and confidentiality of data in the information systems.

KEYWORDS: depersonalization, personal data, hash identifier, hash algorithm, private key, information system.

ABBREVIATIONS

PD is a personal data;

ISPD is an information system of personal data.

NOMENCLATURE

D is a personal data table;

M is a total amount of attributes;

N is a table rows count;

A_1 , A_2 are datasets;

K is a number of key attributes;

F is a hash function;

a_{ik} is a rows of data of the table;

P is an original message;

f is a multi-round non-key reshuffle;

$Theta, Chi, Pi, Rho, Lota$ are hash functions;

A, B, C, D are arrays;

x is an amount;

i is a counter;

Z is a hashing results;

r is an array defining the count of bits of reshuffle for each state;

PK is a private key.

INTRODUCTION

In modern automated systems a large amount of personal data of various security classes is processed. In accordance with the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Strasbourg, January 28, 1981) (with changes and additions), the operator must ensure the confidentiality of the data being processed, which leads to significant material costs [1–3]. So the cost of protecting one workplace of an automated personal data processing system can be more than 1000 US dollars, and the number of workstations of an automated system can be several hundreds of dollars. Also the problem faced by many companies, collecting and storing consents to the processing of personal data that require handwritten completion or using an electronic signature, is known. To solve this problem, the methods of depersonalization can be used [4].

The object of the research is the process of transforming confidential personal data into anonymous, non-confidential sequence.

The process of converting confidential personal data into an impersonal non-confidential sequence usually takes a lot of time, has a low resistance to attacks and has limitations at processing large amounts of personal data with frequent changes.

The subject of the research is the methods of deflating personal data.

Known methods of data depersonalization [5–8] have low speed; in records relationships between attributes of depersonalized data and their corresponding personal data attributes are partially preserved; if the values of individual attributes change, only the composition of the data can change, not the depersonalization. Therefore, in order to increase the speed and confidentiality of data depersonalization, it is necessary to develop a method to eliminate the identified shortcomings.

DOI 10.15588/1607-3274-2020-1-16

The purpose of the work is to increase the speed and quality of the process of depersonalization of data processed in automated information systems.

1 PROBLEM STATEMENT

Let us assume that the raw data is given in a form of preliminary values $D_N(d_1, d_2, ..., d_M)$, where M is the total attribute count and N is the table row count. Attributes d may be key and non-key. As the result the number of the key values is equal to $K\ (0 < K < M)$. While forming of data for hashing a private key PK with bitness of 512 is used. For a given sequence of data, the depersonalization function can be represented as the task of splitting data into two sets, A_1 and A_2, wherein A_1 contains confidential data, A_2 the anonymous information, and finding a unique d_0 sequence such that, for set $F(a_{i1}, a_{i2}, \ldots, a_{ik}, PK)$, the value d_0 will be a unique $d_0 = const$. At the same time the following condition is met – the inverse is impossible, finding a $(a_{i1}, a_{i2}, \ldots, a_{ik}, PK)$ data block from any d_0 is impossible, which in turn allows to establish the interrelation of the elements of the first and second sets.

2 REVIEW OF THE LITERATURE

In the process of analysis of modern methods of PD depersonalization the following methods were studied: method of identifiers implementation, method of change of composition or semantic, method of decomposition, mixing method.

1) Method of identifiers implementation is a replacement of personal data values with creation of a table (guide) of conformity of identifiers with the initial data. The disadvantages of this method are:

a) In the request and in the response to this request the type of representation of PD attributes that were replaced with identifiers is changed.

b) In the records the relations between attributes of depersonalized data and PD attributes corresponding to them are saved.

c) It is applicable to a small amount of PD attributes and the small volume of a PD array.

2) Method of change of composition or semantics is the change of composition or semantics of personal data by replacement with statistic processing, transformation, compilation or replacement of some information [9]. This method has the next disadvantages:

a) Application of this method is uneffective for PD depersonalization, because during PD attributes extracting it is necessary to consider the possibility of depersonalization with the usage of these attributes.

b) During basic replacement of values of separate attributes only change of PD composition can happen, but not depersonalization.

c) In record relations between attributes of depersonalized data and the attributes of personal data corresponding to them are partially saved.

d) Applicable when processing tasks do not require personalization of depersonalized data, if it is needed this process can be used on small data arrays.

3) Method of decomposition is division of an array of personal data into several sub-arrays with subsequent separate storage of sub-arrays . The basic disadvantages are:

a) It saves relations between attributes of depersonalized data and PD attributes corresponding to them in records of each storage.

b) Is applicable on large arrays of PD.

c) Resistance to attacks depends on the complexity of setup of relations between tables

4) Mixing method is a reshuffle of separate values or groups of values of personal data attributed in an array of persona data. This method has these disadvantages:

a) This method does not save relations between attributes of depersonalized data and personal data attributes corresponding to them in records.

b) Resistance to attacks increases with growth of the size of the array of personal data.

c) In applicable to large arrays of personal data with frequent changes in data.

The algorithms for the implementation of the identifiers' priming method are represented by functions, some of which consider various cryptographic approaches for generating an identifier for the connection between the cross-reference table and the depersonalized database. For example, a unique and relevant identifier of an individual is obtained by using a one-way cryptographic function from the following attributes: the surname, name, patronymic and date of birth of the individual – O.A. Vishnyakova and D. N. Lavrov [9]. There is also a patent for a method of identifying a subject of personal data using a SIM card as an identifier for communication, proposed by E. S. Volokitina [10]. The method has been successfully implemented in educational organizations. The featured algorithm successfully solves the security problem during processing anonymous data. However, the use of an additional identifier complicates the processing and increases costs.

Algorithms for the implementation of a method of changing the composition or semantics are presented by I. Y. Kuchin [11], which proposes an approach of encoding identifying attributes based on the developed algorithm. A distinctive feature of the work is the analytical justification for the choice of the composition of the identifying group and the provision of a given degree of anonymity as part of an anonymous database. This method has been introduced in the healthcare field, however, the issue of ensuring security is solved only when storing personal data, not when dealing with other information processing modes.

Algorithms for the implementation of the mixing method are presented by works that propose the use of mixing algorithms aimed at the storage of PI or its transmission over open communication channels. For example, K. O. Bondarenko and V. A. Kozlov [12] have presented a method of mixing data inside segments with sequential

DOI 10.15588/1607-3274-2020-1-16

mixing of rows and sensitive attributes, as the algorithm uses lookup tables generated by the cryptographic gamma method. On the one hand, the use of cryptography guarantees the sustaining power of the algorithm even during a processing session, but, on the other hand, it complicates the process of adding, deleting, searching data and increases the cost of protection. These shortcomings are obstacles for the implementation of the method.

Other research areas involve the use of mainly cryptographic methods, which can be attributed to depersonalization with a sufficient degree of conditionality, since they solve the problem of the impossibility of identifying an individual according to the processed data, but they are not formally included in the set of methods established by Roskomnadzor or merely partially use such methods. For example, the work of Y. V. Trifonova and R. F. Zharinov [13] suggests using the built-in cryptographic tools of the CryptDB database management system. As an example of the partial use of the identifier method, one can cite the work of I. Azhmukhamedov, R. Y. Demina and I. V. Safarov [14], wherein the cross-reference table encryption is applied with subsequent blocking.

To generate a sequence hash, the following method is used based on the concept of a cryptographic sponge, which calls for two primary stages [15–16].

1) Absorbing. The initial message P is subject to multi-round reshuffles *f*, accumulation and processing of all blocks of the message from which the hash will be developed is conducted [17].

2) Squeezing. The output of the received value of Z as the shuffle result, the development of the hash value and the output of the results until the necessary length of the hash is reache [18].

In the absorbing phase first is set the initial state from the zero vector with the size up to 1600 bits . Next is conducted the operation xor of a fragment of the initial message p_0 with the fragment of the initial state with the size of r, the remaining part of the state with capacity of c remains the same.

The result is processed by the f function which is a multiround non-key pseudo-random reshuffle and repeats till the initial message blocks exhaust [19]. Next comes the squeezing phase at which it is possible to extract a hash of a random length. The flow chart of the hashing algorithm is shown at the Fig. 1.

The function $F()$ in this algorithm executes 24 rounds, one round includes the work of five functions *Theta,Chi,Pi,Rho,Lota*, consistently processing the inner state at each round.

The function *Theta* is represented by the next expressions (1):

$$C[x]=A[x,0]\oplus A[x,1]\oplus A[x,2]\oplus A[x,3]\oplus A[x,4], x=0\ ..4$$
$$D[x]=C[x-1]\oplus (C[x+1]>>>1), x=0\ ..4;$$
$$A[x,y]=A[x,y]\oplus D[x], x=0...4, y=0...4. \quad (1)$$

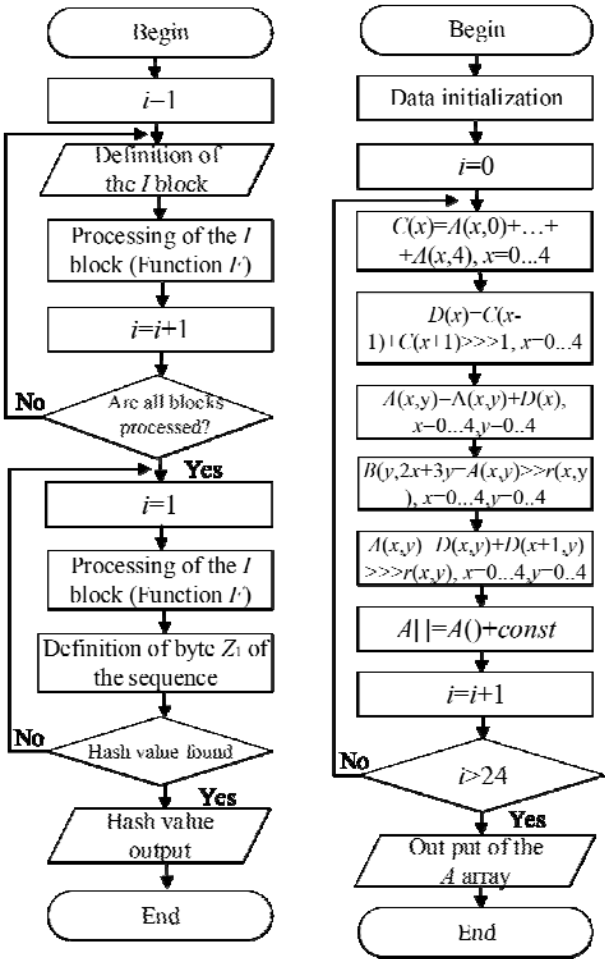

Figure 1 – The flow chart of the hashing algorithm

The function *Chi* is represented by the next expression (2):

$$A[x,y]=B[x,y]\oplus(\sim B[x+1,y]\&B[x+2,y], x=0..4, y=0..4. \quad (2)$$

The functions *Pi*, *Rho* are represented by the next expression (3):

$$B[y,2x+3y]=A[x,y]>>>r(x,y), x=0...4, y=0...4. \quad (3)$$

The function *Lota* is represented by the next expression (4):

$$A[0,0]=A[0,0]\ \text{xor}\ RC. \quad (4)$$

Where B is a temporary array having the same structure as the state array; C and D are the temporary arrays each containing 5 64-bit words; r -array defining the number of bits of spinage for each word of the state; inversion of the value $\sim B[x+1,y]$.

Step1: at the beginning of the algorithm data initialization is conducted. The size of the state is 1600 bits. Next to the variable i the value 0 is assigned.

Step2: after this the processing of the array with functions $C[x]$, $D[x]$, $A[x,y]$, $B[y,2x+3y]$, $A[x,y]$ begins, and besides these operations is conducted the summation of the xor-round constant RC with the word $A[0,0]$.

DOI 10.15588/1607-3274-2020-1-16

Step 3: after data processed with subfunctions goes the check for the rounds count. If the condition $i > 24$ is true then the output of the A array is conducted. If not then we increment by 1 and make the operations until this condition is true.

3 MATERIALS AND METHODS

In order to eliminate the drawbacks mentioned above a personal data depersonalization method, based on the method of identifiers implementation using hashing of critical data and a private key, was developed [20]. As raw data a personal data table $D_N(d_1, d_2, ..., d_M)$ is reviewed, where M is the total amount of attributes and N is table rows count, d_M is an attribute referring to key and non-key.

In this, at the first step by expert way critical data and data clearly identifying the personal data subject is defined. Corresponding attributes are defined as ley ones.

At the second step the initial array D according to chosen key attributes is split into two non-intersecting sub-arrays A_1 and A_2. It is worth noting that into each of sub-arrays an additional attribute d_0 is added, by which value later the comparison of depersonalized data with the personal data subject is conducted. As the result the number of key values is equal to K patients $(0 < K < M)$. . In this, in A_2 is stored depersonalized data that is not interesting for the intruder, so it does not require protection and is stored in the clear.

At the third step for the set of key values of each row $(a_{i1}, a_{i2}, \ldots, a_{in}) \in A_1$, where $i = 1,2,\ldots,N$ the value of the attribute $d_0 = F(a_{i1}, a_{i2}, \ldots, a_{ik}, PK)$ is calculated, where F is a unique function unknown for the user, PK is the unique private key. As F in this case the hash function is chosen [21].

$A_1(d_0, a_1, a_2, \ldots, a_n)$, $A_2(d_0, b_1, b_2, \ldots, b_n)$ where A_1 is the $(a_1, a_2, \ldots, a_n)$ set of confidential data and the d_0 hash, A_2 is the $(b_1, b_2, \ldots, b_n)$ set of anonymous data and the d_0 hash. In addition to the above, knowing the initial $(a_{i1}, a_{i2}, \ldots, a_{in}) \in A_1$ data can contribute to finding the $(b_{i1}, b_{i2}, \ldots, b_{in}) \in A_2$ set.

4 EXPERIMENTS

For the experiments a computer program and a database, implementing the proposed method, with the initial data of 100 subjects of personal data of a medical institution, were developed. The developed software has been studied at solving the problems of depersonalization.

On the basis of the initial sample, key critical attributes were identified that uniquely identify the subject of personal data that is stored in a protected information system. Using this data and a private key, for each record a hash identifier is generated, which is the primary key of the subject of the personal data in the depersonalized information system.

To search for the necessary record in an impersonal information system, a developed subprogram for calculating the identifier hash is used, which based on the data from the primary documents of the personal data subject formed the primary key of the specific record.

After the formation of data for a depersonalized information system, an analysis was performed for the presence of collisions [22–23].

5 RESULTS

As an example let's review a database of patients of some treatment institution (see table 1).

Table 1– Patient database

Last name	First Name	Patronymic	Sex	Date of birth	Medical insurance	Diagnosis
Ivanov	Ivan	Ivanovich	M	12.12.1992	12345678910	Pneumonia
Petrov	Denis	Yurievich	M	11.11.1990	46548677684	Pyelonephritis

For example, for the patient Ivanov the critical personal data is: first name, last name, patronymic, date of birth. For the hash identifier preparation we will use this data:
{Ivanov,Ivan,Ivanovich,12.12.1995}+{bPeShVkYp3s6v9y$B&E)H@McQfTjWnZq}, where the second addend is the private key of the treatment institution. After the calculation we get the hash identifier: 1628b3db5c13865aea5856a630a736653059fc7e2d7c49f897b636428c62a26b.

In the depersonalized database the hash identifier and the depersonalized personal data are stored (see table 2).

Table 2 – Depersonalized database

Hash identifier	Medical insurance	Diagnosis
1628b3db5c13865aea5856a630a736653059fc7e2d7c49f897b636428c62a26b	12345678910	Pneumonia
4d949d630cfaafe3dd151a2e06d7345a44a61889a8c097622abfd6ca0f515a7f	46548677684	Pyelonephritis

In the secure database the hash identifier and the critical personal data are stored (see table 3).

Table 3 – Secure database

Hash identifier	Last name	First name	Patronymic	Date of birth	Hash identifier
1628b3db5c13865aea5856a630a736653059fc7e2d7c49f897b636428c62a26b	Ivanov	Ivan	Ivanovich	12.12.1992	1628b3db5c13865aea5856a630a736653059fc7e2d7c49f897b636428c62a26b
4d949d630cfaafe3dd151a2e06d7345a44a61889a8c097622abfd6ca0f515a7f	Petrov	Denis	Yurievich	11.11.1990	4d949d630cfaafe3dd151a2e06d7345a44a61889a8c097622abfd6ca0f515a7f

DOI 10.15588/1607-3274-2020-1-16

In this, the ability to restore the original data from the hash identifier is impossible. To obtain an identifier it is required to fill in the necessary fields of the subject of personal data from primary documents using the private key in the developed software.

6 DISCUSSION

Let's consider the application of this method sing the famous characters Alice and Bob [24].

Alice came to see Doctor Bob. To identify Alice she shows Bob the critical PD from her initial documents (passport and medical insurance). Bob using the calculator for hash identifier inserts this data and the key of the hospital and forms the hash identifier that allows getting the access to Alice's patient file. After diagnosing and prescribing treatment Bob inserts data into the information system and sings it with his electronic signature.

A curious staff member Eva wanted to know Alice's diagnosis but can't find her card in the information system because she does not know the hash identifier as well as Alice's critical data.

Mallorie found out Alice's critical PD and got the access to the calculator for hash identifier, but she does not know the hospital's key for calculating Alice's identifier.

This method has the next advantages:

1) Data becomes depersonalized which reduces costs of ISPD protection.

2) It is impossible to define the presence of a certain subject in ISPD by known unique attributes.

3) Operator during subject's application by his PD gets access only to one record of ISPD.

4) The context analysis is impossible.

CONCLUSIONS

The actual problem of data depersonalization in the information system was solved by introducing identifiers using hashing of critical data and a private key.

The scientific novelty of the obtained results is that a method was proposed for introducing identifiers using hashing of critical data and a private key for the first time. This allows to increase the level of data confidentiality, reduce the requirements for the level of information system security, increase the speed of data processing by convolving critical data into a hash identifier.

The practical significance of the obtained results is that software that implements the proposed method has been developed and experiments have been carried out to confirm the adequacy of the proposed mathematical model. The results of the experiment allow us to recommend the proposed method for introducing into automated information systems the processing of personal data at the design stage or optimizing of the existing systems, which will reduce the cost of protecting the information system.

Prospects for further research are to explore the possibility of implementing this method in a software and hardware system that allows to increase the speed of the information system.

REFERENCES

1. Rodichev Yu. A. Normativnaya baza i standarty v oblasti informacionnoj bezopasnosti. Sankt-Peterburg, Izdatel'skij dom «Piter», 2018, 255 p.
2. Sychev Yu. V. Standarty informacionnoj bez-opasnosti. Zashchita i obrabotka konfidencial'nyh dokumentov. Saratov, Vuzovskoe obrazovanie, 2019, 223 p.
3. The Convention for the protection of individuals with regard to automatic processing of personal data is a 1981 Council of Europe [Electronic resource]. Access mode: https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680078b37.
4. Regulation (EU) 2016/679 of the European parliament and of the council GDPR(General Data Protection Regulations) [Electronic resource]. Access mode: https://ogdpr.eu/en/gdpr-2016-679.
5. Prikaz Roskomnadzora ot 05.09.2013 № 996 «Ob utverzhdenii trebovanij i metodov po obezlichi-vaniyu personal'nyh dannyh». [Elektronnyj resurs]. Rezhim dostupa: http://www.consultant.ru/document/cons_doc_LAW_151882/
6. Kalutskiy I. V., Shumailova V. A. , Nikulin D. A. et all. Depersonalization of personal data during processing of information in automated systems, *Telecommunications,* 2016, No. 10, pp. 16–20.
7. Spevakova S. V., Primenko D. V. A method of personal data depersonalization in automated systems, *Conference: Optoelectronic devices in pattern recognition systems, image processing and symbol information. Recognition – 2017, Kursk, 16–17 May 2017, proceeding*. Kursk, UZGY, 2017, pp. 330–333.
8. Dobritsa V. P., Gubarev A. A. Algorithm of exclusive transformation of data, *News of the Kursk State Technical University,* 2010, No. 1 (30), pp. 49–54.
9. Vishnyakova O. A., Lavrov D. N. Format obmena dannymi v sisteme sbora i obrabotki biometricheskih obrazcov, *Informacionnye resursy v obrazovanii: mater. mezhdunar. nauch.-prakt. konf*. Nizhnevartovsk, Izdatel'stvo Nizhnevart. gos. un-ta, 2013, pp. 146–149.
10. Volokitina E. S. Metod i algoritmy garantiro-vannogo obezlichivaniya i reidentifikacii sub'ekta personal'nyh dannyh v avtomatizirovannyh informacionnyh sistemah: dis. kand. tekhn. nauk. Sankt-Peterburg, Izdatel'stvo Sankt-Peterburgskogo nac. issled. un-ta informacionnyh tekhnologij, mekhaniki i optiki, 2013, 183 p.
11. Kuchin I. Yu. Obrabotka baz dannyh s personifi-cirovannoj informaciej dlya zadach obezlichivaniya i poiska zakonomernostej: dis. … kand. tekhn. nauk. Astrahan', Izdatel'stvo Astrah. gos. tekhn. un-ta, 2012, 132 p.
12. Bondarenko K. O., Kozlov V. A. Universal'nyj bystrodejstvuyushchij algoritm procedur obezlichivaniya dannyh, *Izv. YuFU. Tekhnicheskie nauki.* Rostov/n/D, Izdatel'stvo YuFU, 2015, No. 11 (172), pp. 130–142.
13. Trifonova Yu. V., Zharinov R. F. Vozmozhnosti obezlichivaniya personal'nyh dannyh v sistemah, ispol'zuyushchih relyacionnye bazy dannyh, *Doklady TUSUR*, 2014, No. 2 (32), pp. 188–194.
14. Azhmuhamedov I. M., Demina R. Yu., Safarov I. V. Sistemnyj podhod k obespecheniyu konfidencial'nosti obezlichennyh personal'nyh dannyh v uchrezhdeniyah zdravoohraneniya, *Sovremennye problemy nauki i obrazovaniya*, 2015, No. 1–1 [Elektronnyj recurs]. Rezhim dostupa: http://www.science-education.ru/ru/article/view?id=18610.
15. Bertoni G., Daemen J., Peeters M., Van G. Keccak code package [Electronic resource]. Access mode: https://github.com/gvanas/KeccakCodePackage
16. /[Huang S., Xu G., Wang M., et all Conditional cube attack on reduced-round Keccak sponge function Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30–May 4, 2017, proceedings. Part II, 2017, P. 259–288.

DOI 10.15588/1607-3274-2020-1-16

17. Guo J., Liu M., Song L. Linear structures: Applications to cryptanalysis of round-reduced Keccak, *International Conference on the Theory and Application of Cryptology and Information Security*. Hanoi, Vietnam, December 4–8, 2016, proceedings. Part I, pp. 249– 274.
18. Jeethu J., Karthikab R., Nandakumarb R. Design and characterization of SHA 3–256 Bit IP core, *International conference on emerging trends in engineering, science and technology, ICETEST*, 2015, Vol. 24, pp. 918–924.
19. Dinur I., Morawiecki P., Pieprzyk J. et all. Cube attacks and cube-attack-like cryptanalysis on the round-reduced Keccak sponge function, *Eurocrypt: Annual International Conference on the Theory and Applications of Cryptographic Techniques Sofia*. Bulgaria, April 26–30, 2015, proceedings, Part I, pp. 733–761.
20. Nozdrina A. A., Spevakov A. G., Primenko D. V.; Patent RF 2636106, MPK G06F 12/14, G06F 12/14. Sposob depersonalizacii personal'nyh dannyh/ zayavitel' Yugo-Zapadnyj gosudarstvennyj universitet. № 2016126867; zayavl. 04.07.2016; opubl. 04.07.2016; Byul. № 32, 4 p.
21. Dobraunig C. Analysis of SHA-512/224 and SHA512/256 / C. Dobraunig, M. Eichlseder, F. Mende // International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29 – December 3, 2015: proceedings. Part I, pp. 612–630.
22. Song L., Liao G., Guo J. Non-full sbox linearization: Applications to collision attacks on round-reduced Keccak, *Annual International Cryptology Conference.* Santa Barbara, CA, USA, August 20–24, 2017, proceedings. Part II, pp. 428–451.
23. Nabeel S., Munqath H. Anti-continuous collisions user based unpredictable iterative password salted hash encryption, *International Journal of Internet Technology and Secured Transactions,* 2018, Vol. 8, No. 4, pp. 619–634.
24. Barakat M., Eder Ch., Hanke T. An Introduction to Cryptography, [Electronic resource]. Access mode: https://www.mathematik.uni-kl.de/~ederc/download/Cryptography.pdf

Received 11.06.2019.
Accepted 23.01.2020.

УДК 004.058.5

МЕТОД ЗНЕОСОБЛЕННЯ ДАНИХ В ЗАХИЩЕНИХ АВТОМАТИЗОВАНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Спеваков О. Г. – канд. техн. наук, доцент кафедри інформаційний безпеки, Південно-Західний державний університет, м. Курськ, Росія.
Спевакова С. В. – аспірант кафедри обчислювальної техніки, Південно-Західний державний університет, м. Курськ, Росія.
Применко Д. В. – аспірант кафедри обчислювальної техніки, Південно-Західний державний університет, м. Курськ, Росія.

АНОТАЦІЯ

Актуальність. Розглянуто завдання знеособлення даних в інформаційних системах. Проведено аналіз сучасних підходів до знеособлювання даних, виявлено та обґрунтовано необхідністю створення нового методу, що дозволяє підвищити захищеність оброблюваних даних і їх достовірність. Об'єктом дослідження є модель деперсоналізації даних, що дозволяє знизити витрати на захист інформаційних систем. Мета роботи – аналіз сучасних методів знеособлення і створення методу, що усуває виявлені недоліки, з підвищеним рівнем конфіденційності та використанням хешування критично важливих даних і приватного ключа.

Мета: аналіз сучасних методів знеособлення і створення методу, що усуває виявлені недоліки, з підвищеним рівнем конфіденційності та використанням хешування критично важливих даних і приватного ключа.

Метод. Запропоновано метод знеособлення персональних даних, заснований на методі введення ідентифікаторів з використанням хешування критично важливих даних і приватного ключа, що дозволяє досягти підвищення конфіденційності інформації, оброблюваної в інформаційних системах. Запропоновано методи вибору ключових критично важливих атрибутів з первинних документів, що дозволяють однозначно ідентифікувати суб'єкта персональних даних, методу формування вихідних множин, розбиває вихідні дані на два непересічних підмножини, методу формування хеш ідентифікатора з унікальної послідовності і приватного ключа, обезлічівающего інформацію і підвищує її конфіденційність.

Результати. Розроблений метод реалізований програмно і досліджений при вирішенні завдань знеособлення.

Висновки. Проведені експерименти підтвердили працездатність запропонованого методу та дозволяють рекомендувати його для впровадження в автоматизованих інформаційних системах обробки персональних даних для вирішення завдань знеособлення. Перспективи подальших досліджень можуть полягати у створенні апаратних засобів потокового знеособлення даних, що дозволяють підвищити швидкість обробки і конфіденційність даних в інформаційних системах.

КЛЮЧОВІ СЛОВА: знеособлення, персональні дані, хеш ідентифікатор, алгоритм хешування, приватний ключ, інформаційна система.

УДК 004.058.5

МЕТОД ОБЕЗЛИЧИВАНИЯ ДАННЫХ В ЗАЩИЩЕННЫХ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Спеваков А. Г. – канд. техн. наук, доцент кафедры информационный безопасности, Юго-Западный государственный университет, г. Курск, Россия.
Спевакова С. В. – аспирант кафедры вычислительной техники, Юго-Западный государственный университет, г. Курск, Россия.
Применко Д. В. – аспирант кафедры вычислительной техники, Юго-Западный государственный университет, г. Курск, Россия.

АННОТАЦИЯ

Актуальность. Рассмотрена задача обезличивания данных в информационных системах. Проведен анализ современных подходов к обезличиванию данных, выявлена и обоснована необходимостью создания нового метода, позволяющего повысить защищенность обрабатываемых данных и их достоверность. Объектом исследования являлась модель деперсонализации данных, позволяющая снизить затраты на защиту информационных систем. Цель работы – анализ современных методов обезличивания и создания метода, устраняющего выявленные недостатки, с повышенным уровнем конфиденциальности и использованием хеширования критически важных данных и приватного ключа.

Цель работы: анализ современных методов обезличивания и создания метода, устраняющего выявленные недостатки, с повышенным уровнем конфиденциальности и использованием хеширования критически важных данных и приватного ключа.

Метод. Предложен метод обезличивания персональных данных, основанный на методе введения идентификаторов с использованием хеширования критически важных данных и приватного ключа, позволяющего добиться повышения конфиденциальности

DOI 10.15588/1607-3274-2020-1-16

информации, обрабатываемой в информационных системах. Предложены методы выбора ключевых критически важных атрибутов из первичных документов, позволяющих однозначно идентифицировать субъекта персональных данных, метода формирования исходных множеств, разбивающий исходные данные на два непересекающихся подмножества, метода формирования хэш идентификатора из уникальной последовательности и приватного ключа, обезличивающего информацию и повышающего её конфиденциальность.

Результаты. Разработанный метод реализован программно и исследован при решении задач обезличивания.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенного метода и позволяют рекомендовать его для внедрения в автоматизированных информационных системах обработки персональных данных для решения задач обезличивания. Перспективы дальнейших исследований могут заключаться в создании аппаратных средств поточного обезличивания данных, позволяющих повысить скорость обработки и конфиденциальность данных в информационных системах.

КЛЮЧЕВЫЕ СЛОВА: обезличивание, персональные данные, хеш идентификатор, алгоритм хеширования, приватный ключ, информационная система.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Родичев Ю. А. Нормативная база и стандарты в области информационной безопасности / Ю. А. Родичев. – Санкт-Петербург : Издательский дом «Питер», 2018. – 255 р.
2. Сычев Ю. В. Стандарты информационной безопасности. Защита и обработка конфиденциальных документов / Ю. В. Сычев. – Саратов : Вузовское образование, 2019. – 223 р.
3. The Convention for the protection of individuals with regard to automatic processing of personal data is a 1981 Council of Europe [Electronic resource]. – Access mode: https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680078b37.
4. Regulation (EU) 2016/679 of the European parliament and of the council GDPR(General Data Protection Regulations) [Electronic resource]. – Access mode: https://ogdpr.eu/en/gdpr-2016-679.
5. Приказ Роскомнадзора от 05.09.2013 № 996 «Об утверждении требований и методов по обезличиванию персональных данных». [Электронный ресурс]. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_151882/.
6. Depersonalization of personal data during processing of information in automated systems/ [I. V. Kalutskiy, V. A. Shumailova, D. A. Nikulin et al.] // Telecommunications. –2016. – № 10. – P. 16–20.
7. Spevakova S. V. A method of personal data depersonalization in automated systems / S. V. Spevakova, D. V. Primenko // Conference: Optoelectronic devices in pattern recognition systems, image processing and symbol information. Recognition – 2017, Kursk, 16–17 may 2017 : proceeding: Kursk, UZGY, 2017. – P. 330–333.
8. Dobritsa V. P. Algorithm of exclusive transformation of data /V. P. Dobritsa, A. A. Gubarev// News of the Kursk State Technical University. – 2010. –№ 1 (30). – P. 49–54.
9. Вишнякова О.А. Формат обмена данными в системе сбора и обработки биометрических образцов / О. А. Вишнякова, Д. Н. Лавров // Информационные ресурсы в образовании: матер. междунар. науч.-практ. конф. – Нижневартовск : Издательство Нижневарт. гос. ун-та, 2013. – С. 146–149.
10. Волокитина Е. С. Метод и алгоритмы гарантированного обезличивания и реидентификации субъекта персональных данных в автоматизированных информационных системах: дис. канд. техн. наук / Е. С. Волокитина. – СПб. : Издательство Санкт-Петербургского нац. исслед. ун-та информационных технологий, механики и оптики, 2013. – 183 с.
11. Кучин И. Ю. Обработка баз данных с персонифицированной информацией для задач обезличивания и поиска закономерностей: дис. … канд. техн. наук / И. Ю. Кучин. – Астрахань : Издательство Астрах. гос. техн. ун-та, 2012. – 132 с.
12. Бондаренко К. О. Универсальный быстродействующий алгоритм процедур обезличивания данных / К. О. Бондаренко, В. А. Козлов // Изв. ЮФУ. Технические науки. – Ростов/н/Д: Издательство ЮФУ. – 2015. – № 11 (172). – С. 130–142.
13. Трифонова Ю. В. Возможности обезличивания персональных данных в системах, использующих реляционные базы данных / Ю. В. Трифонова, Р. Ф. Жаринов // Доклады ТУСУР. – 2014. – № 2 (32). – С. 188–194.
14. Ажмухамедов И. М. Системный подход к обеспечению конфиденциальности обезличенных персональных данных в учреждениях здравоохранения / И. М. Ажмухамедов, Р. Ю. Демина, И. В. Сафаров // Современные проблемы науки и образования. – 2015. – № 1–1 [Электронный ресурс]. – Режим доступа: http://www.science-education.ru/ru/article/view?id=18610.
15. Keccak code package [Electronic resource] / [G. Bertoni, J. Daemen, M. Peeters, G. Van]. – Access mode: https://github.com/gvanas/KeccakCodePackage
16. Conditional cube attack on reduced-round Keccak sponge function / [S. Huang, G. Xu, M. Wang et all.] // Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 – May 4, 2017: proceedings. Part II, 2017. –P. 259–288.
17. Guo J. Linear structures: Applications to cryptanalysis of round-reduced Keccak / J. Guo, M. Liu, L. Song // International Conference on the Theory and Application of Cryptology and Information Security, Hanoi, Vietnam, December 4–8, 2016: proceedings. Part I. –P. 249– 274.
18. Jeethu J. Design and characterization of SHA 3– 256 Bit IP core / J. Jeethu, R. Karthikab, R. Nandakumarb // International conference on emerging trends in engineering, science and technology, ICETEST. – 2015. – Vol. 24. –P. 918–924.
19. Cube attacks and cube-attack-like cryptanalysis on the round-reduced Keccak sponge function / [I. Dinur, P. Morawiecki, J. Pieprzyk et al.] // Eurocrypt: Annual International Conference on the Theory and Applications of Cryptographic Techniques Sofia, Bulgaria, April 26–30, 2015: proceedings. Part I. – P. 733–761.
20. Патент РФ 2636106, МПК G06F 12/14, G06F 12/14. Способ деперсонализации персональных данных/ А. А. Ноздрина, А. Г. Спеваков, Д. В. Применко; заявитель Юго-Западный государственный университет. – № 2016126867; заявл. 04.07.2016; опубл. 04.07.2016; Бюл. № 32. – 4 с.
21. Dobraunig C. Analysis of SHA-512/224 and SHA512/256 / C. Dobraunig, M. Eichlseder, F. Mende // International Conference on the Theory and Application of Cryptology and Information Security, Auckland. – New Zealand, November 29 – December 3, 2015: proceedings. Part I. –P. 612–630.
22. Song L. Non-full sbox linearization: Applications to collision attacks on round-reduced Keccak / L. Song, G. Liao, J. Guo // Annual International Cryptology Conference. – Santa Barbara, CA, USA, August 20–24, 2017: proceedings. Part II. – P. 428–451.
23. Nabeel S. Anti-continuous collisions user based unpredictable iterative password salted hash encryption / S. Nabeel, H. Munqath // International Journal of Internet Technology and Secured Transactions. – 2018. – Vol. 8, № 4. –P. 619–634.
24. Barakat M. An Introduction to Cryptography, [Electronic resource] / M. Barakat, Ch. Eder, T. Hanke. – Access mode: https://www.mathematik.uni-kl.de/~ederc/download/Cryptography.pdf

DOI 10.15588/1607-3274-2020-1-16