Webology (ISSN: 1735-188X)
Volume 19, Number 2, 2022

Reliability Improvement Of Communication Channels Between The Components Of Distributed Information Systems

Tanygin Maxim Olegovich[1] , Ali Ayid Ahmad[2] , Dobritsa Vyacheslav Porfirevich[3] , Huseyin POLAT[4] , Ahmad Ayid Ahmad[5]

[1]Ph.D. Associate Professor, Head of the department of Information Security, Southwest State University, St. 50 years of October, 94. Kursk, The Russian Federation. ORCID: 0000-0002-4099-1414.

[2]Post-graduate student of the department of Information Security, Southwest State University from Iraq. Kirkuk University, St. 50 years of October, 94. Kursk, The Russian Federation. ORCID: 0000-0002-6031-9414.

[3]Dr. of. Sci., Professor, Department of Information Security, Southwest State University St. 50 years of October, 94. Kursk, The Russian Federation. ORCID: 0000−0001−7533−3684.

[4]Ph.D. Associate Professor, Faculty of Technology, Department of Computer Engineering, Gazi University, 6/19 Bandırma St, Yenimahalle, ANKARA, 06560, TURKEY. ORCID: 0000-0003-4128-2625.

[5]Ph.D. Student, Faculty of Technology, Department of Computer Engineering, Gazi University, from Iraq. Kirkuk University, 6/19 Bandırma St, Yenimahalle, ANKARA, 06560, TURKEY.

Abstract

A particular class of communication protocols has been created to ensure the interaction of specialized process control devices and complex technical objects. Several methods based on determining the source for groups of messages have been created to increase the reliability of identification. However, they suffer from high computational complexity, which depends on the number of possible options for forming such groups from the entire set of messages processed by the receiver. This study aims to increase the reliability of identification and reduce the computational

complexity of existing algorithms. Therefore, this article proposes a method of limiting the set of data packets processed by the device. The method can be used in protocols in which the size of the identification information field is limited to a few bits. The evaluation demonstrates that limiting the number of messages and the occurrence reduces 1.5 to 3.0 times the probability of an error when defining the source of message groups. Furthermore, the frequency of repeated questions caused by identification errors is reduced, which ensures a reduction in the information redundancy of the transmitted data and an increase in the speed of their processing.

Keywords: message flow processing, information security, identification, mathematical modeling, identification error.

Introduction

In modern distributed information systems, the reliability of processing data transmitted between individual components within a group of computer systems determines the functioning quality of the distributed information. This reliability can only be ensured by identifying the source of each data packet transmitted among such systems [1].

The challenge of recognising a distant subject can be expressed in general as follows: A data receiver, which receives a specific set of U messages across the communication channel. The receiver communicates with a variety of I, II, III, and other sources. It is important to choose disjoint subsets U_I, U_{II}, U_{III}, etc. from the incoming set of data packets U, which were generated and transmitted by the corresponding sources to the receiver [2].

There are many approaches for detecting the source of each message received by the receiver, but they all rely on encrypting a digital fingerprint from the payload of the u^{inf}c message with the secret key Ω [3]. The encryption result $\boldsymbol{F}\left(u^{\inf},\Omega\right)$, in this case, is inserted into the message and sent together with the payload:

$$\begin{gathered}\boldsymbol{G}\left(u,\Omega_{\mathrm{I}}\right)=1\Rightarrow u\in U_{\mathrm{I}},\boldsymbol{G}\left(u,\Omega_{\mathrm{I}}\right)=0\Rightarrow u\notin U_{\mathrm{I}},\\ \boldsymbol{G}\left(u,\Omega_{\mathrm{II}}\right)=1\Rightarrow u\in U_{\mathrm{II}},\boldsymbol{G}\left(u,\Omega_{\mathrm{II}}\right)=0\Rightarrow u\notin U_{\mathrm{II}},\\ \ldots\end{gathered} \qquad (1)$$

where **G** is the deciding function for the incoming message u and the source identifier Ω_I, allowing you to choose a subset of U_I from the set U. Because the receiver has no prior knowledge of the contents of the field u^{inf} the decision is made solely on the basis of the value analysis of word $\boldsymbol{F}\left(u^{\inf},\Omega\right)$. This word usually called message authentication code (MAC) [4]. The following is an example of an erroneous condition:

$$u\notin U_A\wedge\boldsymbol{G}\left(u,\Omega_A\right)=1,u=\left\{u^{\inf},\boldsymbol{F}\left(u^{\inf},\Omega\right)\right\}. \qquad (2)$$

where A is the target source of messages [5].

As a result, the reliability of identification can be estimated by the ratio of two values: the power of the set U and the number of code combinations of the MAC $\boldsymbol{F}\left(u^{\inf},\Omega\right)$ which is given by the bit depth (H) of the processed word [6]. For all words in the set U that do not belong to the source A, the probability of a successful message source identification is defined as the probability of non-fulfillment of condition (2) [7]:

$$p_{\text{tr}}^{1}=\left(1-2^{-H}\right)^{|U|}\approx 1-2^{-H}\,|U| \text{ while } 2^{H}>|U|. \tag{3}$$

Protocol identification approaches focus on the communication channel's limited bandwidth [8], and the size of the transmitted message can be reduced to a few tens or even a few bytes [9]. As a result, the MAC size prevents components of distributed systems from being reliably identified. The approaches used to boost the reliability of identification are based on the decisive rule application to a group of messages rather than an individual message [10]:

$$\begin{aligned}&\boldsymbol{G}\left(\{u_1,\ldots u_M\},\Omega_{\text{I}}\right)=1\Rightarrow u_1\in U_{\text{I}},\ldots,u_M\in U,\\&\boldsymbol{G}\left(\{u_1,\ldots u_M\},\Omega_{\text{I}}\right)=0\Rightarrow u\notin U_{\text{I}},\ldots,u_M\notin U,\end{aligned} \tag{4}$$

where M denotes the size of the message group for which the belonging to the set is determined. The probability of successfully identifying the source of messages for them is higher than the probability of successfully identifying the source of M messages using single-message analysis methods [11]:

$$p_{\text{tr}}^{M}=\left(1-2^{-H\cdot M}\right)^{A_{|U|}^{M}}>\left(p_{\text{tr}}^{1}\right)^{M}. \tag{5}$$

where $A_{|U|}^{M}$ is the number of combinations M by |U|, which corresponds to the number of alternatives for building a group of M messages from the set U (we assume the order of messages in condition (4) is crucial for the outcome of applying the rule **G**).

RELATED WORKT

he problem of assembly and verification of data thus obtained is dedicated to the study [12]. Here the error-free identification of attributes of each block is determined by the results of operations in the Galois fields over two matrices: one, a check matrix, which is stored in receiver and sender, and a second one, which is formed by vectors - values of obtained blocks. For transformations, the check matrix is augmented by pseudorandom numbers. Based on the structure of matrix obtained as a result of operations on the above two matrices: a verification matrix and a vector of received messages, a decision on the correctness of information extraction about the source of the information block being processed and its position in the final fragmented message is made. The method allows providing reliability and inaccuracy of processing at sender terminals of the information divided into separate fragments due to the universal numerical characteristic describing relative arrangement of blocks

relative to each other, with simultaneous accounting of their contents. This method is based on the principle of "joint correctness" - when reliability of its attributive information determination is estimated not in isolation from other sequence blocks, but by its belonging to a structured set. This affiliation is determined by analysis of a generalized formalized result that is obtained from analysis of the whole set of blocks. At the same time, this method only controls the attributive information determination correctness, without the ability to control the determination process and dynamically adjust receiver working modes to achieve the required reliability parameters. If we talk about the complexity of the algorithm, in the case where there are m data blocks from which it is required to form a message of length n (n<m), the number of variants of the input vector when checking the placement of blocks in it is determined by the number of permutations n by m. That at n<m makes the number of required calculations very large: $O\left(A_m^n\right)$ [13], which is limited its applicability to sequences of small length only and reduces the reliability (5).

Simultaneously, the described principle of "joint correctness" is acceptable for determining the correctness of the position of parts of messages in the stream, whose size does not allow to place in them all the attributive information: the source, the place of this part in the message, and the data integrity flag [14].

A similar approach to the above is the information block sequence control proposed by Prem kumar and Chant. It is based on matrix transformations at the individual block level, with the goal of increasing the rate of computation of some numerical characteristic for data blocks [15]. In their approach, the data is partitioned into several blocks and combined into a square matrix. The determinant of each matrix is generated and can be used for multiple purposes, including identifying the source that generated the information block. This numerical sequence is then encoded using a combination of reversible and irreversible transformations. The data-processing device sequentially rearranges the resulting blocks to produce an initial arrangement that will give a match to some known numerical factor known in advance. It should be noted that the method under consideration, in which the principle of joint correctness is also implemented, is focused more on ensuring the impossibility of restoring the structure of the data matrix without a known digital factor, rather than on providing a quick check of the IS for their belonging to the source. The approach based on the formation of the data matrix from individual blocks has a factor complexity of O(n!) for reconstruction of the original sequence at the receiver side, which limits its application to identify the source of messages from numerous blocks received from a large number of sources.

In the consider protocols the size of a data block does not exceed several tens or even several bytes. In the system described in [16], the H/(L+H) ratio reaches 30% due to the use of additional fields to allocate data block attributes of small size and the independent execution of such allocation from others for each IS. In this regard, the approach used in [12], based on the group verification of blocks coming into the receiver, seems more effective in terms of information redundancy. The disadvantage of reducing the information redundancy is the need to re-transmit and process the entire group of blocks, processed with an error. This, as mentioned above, increases the duration of the full cycle of

message processing indirectly, due to the need to re-transmit the entire set of data blocks. Therefore, the method of group control of the source and structure blocks group must be combined with methods that reduce the probability of errors that require retransmission.

Separately, it is worth mentioning the additional time costs that arise when using methods to control delivery route correctness. In the considered work [16] it has a linear dependence on the number of blocks received by the receiver per unit time, which in itself is a good result. In [12], the same additional time costs are required to multiply the matrices, the size of which is proportional to the number of message parts. Therefore, the complexity of this procedure is proportional to the square of the number of parts of one message. In the same way, the memory requirements for buffering the fragmented message grows quadratic ally.

Everything described above causes the necessity to use the algorithms excluding change of sequence of data blocks, including the algorithms based on integration of blocks into structures [17], in which attributes of an individual block are determined by its belonging to such a single chain and by its location in it [18].

Evidently, incorrect processing of a chain distorted by inclusion of one or more extraneous blocks into its composition, even if it consists of ten IB, is less probable [19]. The task of the formulated methods of reducing the information redundancy is to exclude the embedding of extraneous blocks into the processed groups of blocks. In this case, the term "extraneous block" in this paper is understood as a block, issued not by the target source, whose data are analyzed by receiver.

The high complexity of the methods for identifying the source of message groups is a disadvantage, because in order to apply the decisive rule , $A_{|U|}^{M}$ verifiable sets are usually required, as in work [10] and [11]. As a result, methods for the arbitrary creation of a group of checked messages are utilized for the possibility of implementation, rather than methods for directed search for choices for arranging messages in a group. A variety of approaches for generating messages were considered $u=\left\{u^{\inf},\boldsymbol{F}\left(u^{\inf},\Omega\right)\right\}$ in the work [20], which, under certain circumstances, produced a quadratic or even linear complexity of the size implementation of the set U, providing a confidence greater than a given inequality (3).

Based on the abovementioned, the main approach for improving the reliability of identifying the source of a message while also minimizing the complexity of implementing this procedure is to minimise the size of the set U of messages handled by the receiver through applying extra constraints.

Problem Statement

Based on the methods for determining the source of information mentioned in the literature, a method was created that consisted of building a source group U_A from M messages. In addition to the word $\boldsymbol{F}\left(u^{\inf},\Omega\right)$, the index J of the message in the group is entered in each message, and taking values

that range from 1 to M, respectively [21]. In comparison to existing approaches, the present method allows for substantially reduced computational complexity [22], with maintaining the length of the encryption field:

$$H < \log_2\left(\frac{|U|}{M}\right). \quad (6)$$

The likelihood of error, on the other hand, increased significantly, rendering the method unsuitable for practical applications. It's vital to formulate a rule that generates a smaller set of analysed messages U from the set U of messages received by the receiver during the transmission of the set U'. As a result, the probability of correctly identifying the message source will increase:

$$\left(1-2^{-H\cdot M}\right)^{A^M_{|U|}} > \left(1-2^{-H\cdot M}\right)^{A^M_{|U'|}}, |U| > |U'|. \quad (7)$$

Materials and methods

The order of development and output of data blocks composing the set U_A provides the theoretical foundation for the formulation of the rule that restricts the power of the set U'. Because the set U_A messages are constructed and transmitted consecutively from the first to the mth, it is reasonable to suppose that some variations in the sequence in which blocks are received by the receiver are not possible. The first block, for example, may come after the second, but not after the third.

As the J Index J value is available in each message, it can be divided into M disjoint subsets w1 – wM, each of which contains only one message from the set U_A and an unlimited number of messages from the set U/U_A.

Considering a random moment in the transmission time of a bunch of messages of the set U_A, let M_{max} be the received message's maximum index. If the condition is fulfilled, the following message u will belong to the set U' when it arrives at the receiver:

$$M_{\max} + W_{\text{back}} \le J^u \le M_{\max} + W_{\text{forw}} \quad (8)$$

where J^u denotes the index of the received message u, and W_{forw} is the advance window's width, the maximum value by which the index of the incoming data block can exceed the maximum index of messages M_{max}, $1 \le W_{forw} \le M$.

W_{back} represents the delay window's width, a parameter that specifies the maximum value by which the incoming message's index can be less than $M_{max} \le 1$ $W_{back} \le M$.

As a result, the power of the analysed messages of the set |U'| will be as max (M, W_{back} + W_{forw})/M times less than the power of the set U, allowing evaluation of the effectiveness of the method by knowing that successful identification of probability dependence p_{tr}^M varies with the change in the

parameters H, |U'|, and M (5). The dependence of the presently studied identification method p_{tr}^{M} = f(H,|U'|,M), which was adopted by [22] through presenting the analysis process of the message as a linear dynamic system.

At the same time, the set formation concept allows for cases where the messages of the set U_A do not fit into the set U'. A mathematical model of receiving messages by the receiver is developed to evaluate the probability of this occurrence, based on the definition of Markov Process of receiving messages with continuous time. For a unit of model time, the time of transmission of a set of U_A messages is selected. The graph assumes the presence of several absorption states: One relates to the hit of all messages from U_A to U' ($U_A \cap U' = U_A$), while the others relate to the non-occurrence of some messages from U_A in U' ($U_A \cap U' \neq U_A$). After a sufficiently enough period of time, the probability of the system entering the first state corresponds to the probability of the P_U successfully forming the set U' (Figure 1).

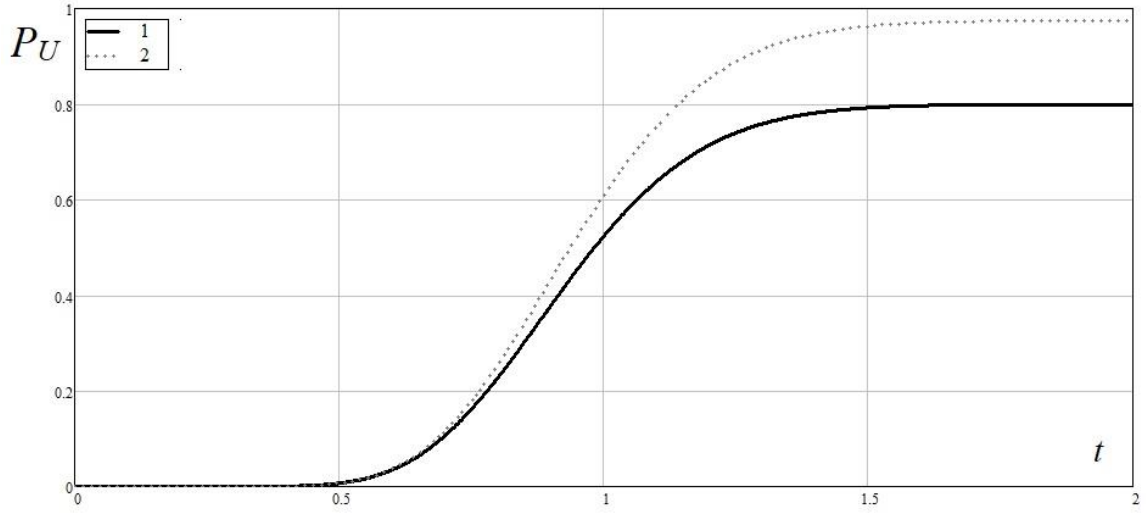

Fig. 1 Depicts the dependence of the probability of successful development of the set U' on the conditional time.

1) H = 6, |U| = 70, M = 20, W_{back} = 10; 2) H = 6, |U| = 70, M = 20, W_{back} = 12;

The simulation revealed that changing the value of W_{forw} has no effect on the probability of P_U, but only enhances the power of the set U'. This finding and the following results are presented for the value W_{forw} = 1.

Given that an identification error will occur either if some of the messages from U_A do not fit in U', or if $U_A \cap U' = U_A$, condition (3) is fulfilled for more than one set of messages with power M. The overall probability of an error is accordingly evaluated as follows:

$$P_{err} = 1 - P_U + P_U (1 - p_{tr}^{M}) = 1 - P_U \, p_{tr}^{M} \quad (9)$$

Results

Using the method of limiting the number of analysed messages illustrates by the graphs in Figure 2, where the dependence between identification error probability and the width of the W_{back} delay window, the size of the message group M, the number of messages in the set U, and the size of MAC H is shown. The same graphs of the dependence of the probability of an identification error showed comparable pattern when analysing the whole set of U messages. Furthermore, both graphs showed coincidence in the range of values where $W_{back} \geq M$, as in this case the cardinalities of the sets U and U' are equivalent.

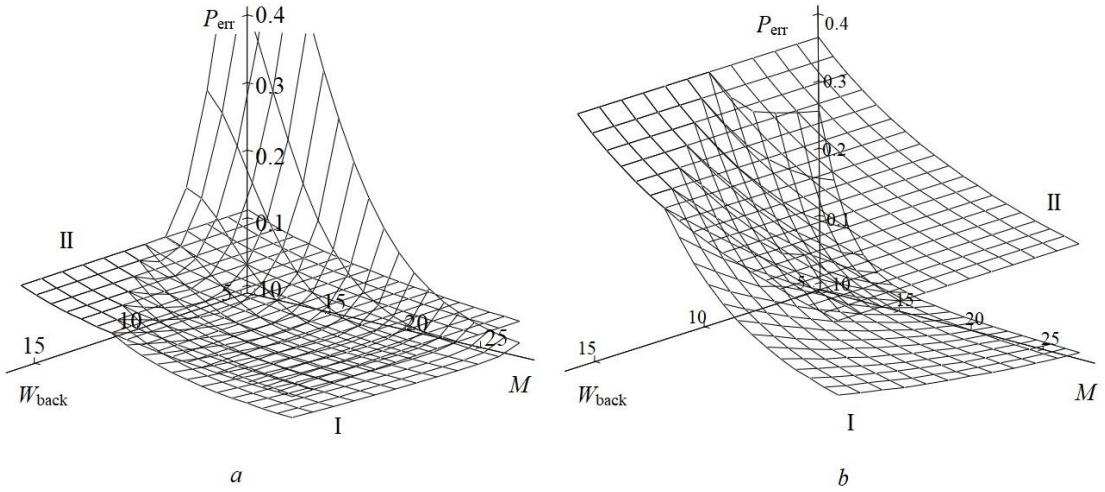

Fig. 2 shows the dependence of the probability of identification error at |U|=150:

I – using the method of limiting the number of analyzed messages,

II – without using the method of limiting the number of analysed messages,

a) H = 7 b) H = 6.

Figure 2b represents the probability of an identification error in this range of parameters without applying the method of limiting the number of analyzed messages reaching values of 0.2 to 0.4. While Figure 2a shows the probability of an identification error smaller than 0.15 at the start. Therefore, a large contribution $1-p_{tr}^{M}$ of the probability of an error of non-occurrence in the U' part of U_A messages, is evident and may surpasses this value, but reduces at the application range of the method (W_{back} and M values for which $P_{err} < 1-p_{tr}^{M}$). However, the ratio between error probabilities in this range is P_{err} and $1-p_{tr}^{M}$ stays at 1.5 to 3.0.

Conclusion

The usage of message group analysis methods is the only option to provide the precise level of identification reliability when both, transmitted messages and identifiers transmitted in such

messages have limited lengths. In this situation, the identification reliability of the methods contradicts with the computational complexity of the processing steps for evaluating the source of messages.

In the known approaches described in the literature, the complexity of message extraction from a required source grows factorially with increasing number of the total number of messages analyzed. When authentication is performed for multiple message sources, arriving at the receiver in random order, the computational complexity of message group processing operations increases dramatically. At the same time decreases the reliability of authentication due to the probability of coincidence of the MAC in messages from different sources. In this paper we consider an approach based on the stability of the characteristics of the flow of messages from source to receiver. This allows the procedures for processing groups of messages to reduce the number of different options for the formation of such groups and increase the reliability of authentication. The created mathematical models of message flow based on Markov chains theory allowed to estimate the resulting effect of applying the method of limiting the number of processed messages

The present study demonstrates the effect of limiting the number of messages and the occurring reduction, by 1.5 to 3.0 times, of the probability of an error when defining the source of message groups. Considering the complexity and reliability of probability identification procedures, future research in this area can be designed to formulate a complicated target and characteristic process of receivers of limited-length messages.

Studies have shown that the maximum efficiency of the proposed method of limiting the number of messages to be processed is when the MAC size does not allow authentication with a confidence higher than 0.8

	Probability of authentication error	Complexity	MAC size in bits
A method based on processing all incoming messages to the receiver	$1-\left(1-2^{-H\cdot M}\right)^{A^M_{\lvert U\rvert}}$	$O(A^M_{\lvert U\rvert})$	H >7
A method based on processing a limited set of messages	$0.3\ldots0.5\times\left[1-\left(1-2^{-H\cdot M}\right)^{A^M_{\lvert U\rvert}}\right]$	$O(A^M_{0.2\lvert U\rvert}) \ldots O(A^M_{0.5\lvert U\rvert})$	H > 5

Table 1: Comparison of the characteristics of the proposed method and methods based on the processing of all incoming messages

Based on the outcomes of the further research in the area of receiver process parameters, the domains of maximum conducted values and derivatives, can then be evaluated. Process parameters represent the theoretical foundation for the development of protocols and devices of advanced operational properties in the situations of modifying communication channel characteristics.

References

1. Stallings, W. (2010). NIST block cipher modes of operation for authentication and combined confidentiality and authentication. Cryptologia, 34(3), 225-235.

2. Dworkin, M. J. (2007). Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC.

3. Burda, K. (2006). Error propagation in various cipher block modes. IJCSNS, 6(11), 235.

4. Black, J., & Rogaway, P. (2000, August). CBC MACs for arbitrary-length messages: The three-key constructions. In Annual International Cryptology Conference (pp. 197-215). Springer, Berlin, Heidelberg.

5. Bellare, M., & Namprempre, C. (2008). Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. Journal of cryptology, 21(4), 469-491.

6. Dworkin, M. (2003). Block Cipher Modes of Operation: The CCM Mode for Authentication and Confidentiality. NIST Special Publication, 800, 38C.

7. Bogdanov, A., Khovratovich, D., & Rech berger, C. (2011, December). Biclique cryptanalysis of the full AES. In International conference on the theory and application of cryptology and information security (pp. 344-371). Springer, Berlin, Heidelberg.

8. Vangelista, L. (2017). Frequency shift chirp modulation: The LoRa modulation. IEEE Signal Processing Letters, 24(12), 1818-1821.

9 Goursaud, C., & Gorce, J. M. (2015). Dedicated networks for IoT: PHY/MAC state of the art and challenges. EAI endorsed transactions on Internet of Things.

10. Black, J., & Rogaway, P. (2002, April). A block-cipher mode of operation for parallelizable message authentication. In International Conference on the Theory and Applications of Cryptographic Techniques (pp. 384-397). Springer, Berlin, Heidelberg..

11 Black, J., Halevi, S., Krawczyk, H., Krovetz, T., & Rogaway, P. (1999, August). UMAC: Fast and secure message authentication. In Annual International Cryptology Conference (pp. 216-233). Springer, Berlin, Heidelberg.

12. Papadimitratos, P., & Haas, Z. J. (2003). Secure message transmission in mobile ad hoc networks. Ad Hoc Networks, 1(1), 193-209.

13. Roman'kov, V. (2020). Algebraic cryptanalysis and new security enhancements. Moscow Journal of Combinatorics and Number Theory, 9(2), 123-146.

14. Shi, X., & Xiao, D. (2013). A reversible watermarking authentication scheme for wireless sensor networks. Information Sciences, 240, 173-183.

15. Prem kumar, P., & Shanthi, D. (2016). Block Level Data Integrity Assurance Using Matrix Dialing Method towards High Performance Data Security on Cloud Storage. Circuits and Systems, 7(11), 3626.

16. Othman, S. B., Alzaid, H., Trad, A., & Youssef, H. (2013, July). An efficient secure data aggregation scheme for wireless sensor networks. In IISA 2013 (pp. 1-4). IEEE.

17. Bista, R., Jo, K. J., & Chang, J. W. (2009, December). A new approach to secure aggregation of private data in wireless sensor networks. In 2009 eighth IEEE international conference on dependable, autonomic and secure computing (pp. 394-399). IEEE.

18. Kaul, V., Bharadi, V. A., Choudhari, P., Shah, D., & Narayankhedkar, S. K. (2015, February). Security enhancement for data transmission in 3G/4G networks. In 2015 International Conference on Computing Communication Control and Automation (pp. 95-102). IEEE.

19. Sumathi, R., Kirubakaran, E., & Thangavel, M. (2012, December). A secure data transfer mechanism using single-handed re-encryption technique. In 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET) (pp. 1-9). IEEE.

20. Tanygin, M. O., Alshaia, H. Ya., & Dobrica, V. P. (2020). Evaluation of the influence of the organization of buffer memory on the speed of execution of procedures for determining the source of messages. Proceedings of the MAI, (114), 14-14.

21. Tanygin, M. O., Alshaeaa, H. Y., & Kuleshova, E. A. (2020). A method of the transmitted blocks information integrity control. Radio Electronics, Computer Science, Control, (1), 181-189.

22. Tanygin, M. O., Dobritsa, V. P., & Alshaeaa, H. Y. (2020, September). Study of the Influence of the Unauthorized Blocks Number on the Collision Probability. In International Russian Automation Conference (pp. 111-120). Springer, Cham.