Study of the influence of the unauthorized blocks number on the speed and RAM expenses during the data analysis process

*Tanygin Maxim Olegovich*
Department of Information Security, Faculty of Fundamental and Applied Informatics, Southwest State University.
Kursk, The Russian Federation.
tanygin@yandex.ru

*Dobritsa Vyacheslav Porfirevich*
Department of Information Security, Faculty of Fundamental and Applied Informatics, Southwest State University.
Kursk, The Russian Federation.
dobritsa@mail.ru

*Hyder Yahya Alshaeaa*
Department of Computer Sciences, Faculty of Education for Pure Sciences, University of Thi-Qar.
Thi-Qar, Iraq.
haideryhy7@gmail.com

**Abstract**—In order to the proper operation of the hardware and software systems and for increase of the reliability of legal data, it is necessary for the hardware device to receive data only from the corresponding software. Otherwise, the data received from extraneous device can lead to errors in the operation of the device or even to a complete loss of its functionality or data. In order to identify the challenges of the transfer of blocks, this study focuses on the influence of the number of unauthorized blocks, on the speed and RAM expenses during the data analysis process. The described method allows to reduce the costs of hardware, which exchanges and analyzes blocks of information. This is done with the help of a buffer to store information and with a set of mathematical equations. And measure the extent of the effect of intensity of receiving unauthorized blocks and hash field length.

**Keywords**—Unauthorized Blocks; Method of reducing hardware costs; Hash; Buffer; Analysis of information blocks.

1. **INTRODUCTION**

During the rapid development of the Internet, the integrity of received information and its reliability play an invaluable role. The traditional method of transmitting information between the source and receiver devices is the use of protocols, where the most modern block-by-frame transmission protocols provide procedures for organizing received frames [1] [2], which divide the data stream into several packets [3]. The typical structure of the information package as shown in Fig.1 [4].

![Figure 1. — The typical structure of the information package (frame).](image-url)
Any data packet passed between the source and receiver contains additional fields or metadata in addition to the service information fields as shown in Figure 1. Based on this data, the receiver determines the sender, the structure of the message formed from a set of information packets and the other transmission parameters [5] [6] [7]. At the same time, if the physical conditions of transmission impose serious restrictions on the size of the data frame transmitted per session, the size of such fields creates a fairly significant information redundancy [8] [9].

If we assume that the transmitted information is divided into a set of information blocks, we can use a method based on the belonging of such separate parts to a single logically related structure, but it is not guaranteed that the information will be received by the receiver in the same order. There can be several reasons for this. First of all, these are the features of the routing of information packets, if transmission occurs through a switched channel [10]. In the case, if the communication channel cannot be switched, for example, in radio communication, transmission errors may occur [11]. It requires the re-transmission of a specific information block in the sequence. A kind of non-switched channel, which sometimes requires re-transmission of information blocks, which is called map-oriented communication protocols [12], such as Peripheral Component Interconnect Express PCI-Express [13], which is focused on high-speed data transmission without loss. In this regard, the most modern block-by-block (frame-by-frame) transmission protocols provide procedures to streamline the received frames.

In addition, we consider a typical system for transmitting information blocks IB (packets) with multiple sources and a single receiver [14]. The solution of these issues can be the use of special algorithms that can control the order and authenticity of information blocks "on top" of the Protocol that provides the transmission. This paper focuses on the analysis of the characteristics of an algorithm that provides secrecy and authenticity of the small data block size (a few tens of bits) and to increase the speed of analysis of service data. The limited size of the processed information and output words makes it impossible to use well-known and well proven irreversible (cryptography hashing) and reversible (encryption) algorithms. Consequently, we require original algorithms that would give satisfactory results on the entropy of output words on small sizes of the initial data [15], and reduce the time and material cost of developing methods analysis of service data.

2. METHOD OF DATA CONTROL

To reduce the probability of errors in the transmission of small data blocks, we advise to use algorithms based on the formation of chains of related information blocks, in which the integrity and authenticity of a single block are determined by its belonging to such a single chain [16]. A similar approach is used in the well-known block chain technology [17]. Obviously, it is more difficult to forge the entire chain, even consisting of a dozen packets, than to falsify one or two transmitted packets. Therefore, for an attacker, the easiest way to influence the described systems is to form random blocks and send them to the receiver [18]. The receiver, when forming a chain of blocks, may include one or more similar unauthorized block in it by mistake [18]. As a result, a chain of information packets from a legal source will be transmitted with an error and will not be processed, which is an implementation of a well-known type of denial-of-service (DoS) attacks [20]. Therefore, it is important to protect the data transferred between sender and receiver [21]. The algorithm of forming data blocks and their subsequent integration into chains is described in detail in the works [22] [23].

The data exchange Protocol described in detail in [24] is based on the adding in the format of the transmitted data block $S_i$ special fields containing:

- information $S_{index}$ about the serial block number $i$ in the chain $s_{index} = F_{crypt} (s^{sec}_i)$, where $F_{crypt}$ – cryptography transformation, $S^{sec}$ – a secret key, known only to the receiver and the sender of the messages.
- hash sequence formed from the information contained in the previous block of the chain $S^{hash}_i = F_{hash} (s^{inf}_{i-1})$, where $F_{hash}$ is a function of cryptography hashing of several words and formation of a simulated device from an arbitrary amount of data of a given size, $s^{inf}_{i-1}$ – information part of the previous block of the chain. The transmitted block itself can be written as: $S = \{S_{inf}|S_{index}|S^{hash}\}$.

3. RESOURCE COMPLEXITY IN THE BUFFER

The formation of structured sequences of information blocks requires the analysis of a set of blocks arriving to the device. If the information is divided into a certain set of information blocks $\{s_1, \ldots, s_n\}$, to control the source we use a method based on the belonging of such individual parts to a single logically related structure, as shown in Fig.2.

where: $U$ is a set of information blocks received by the receiver, with $\tilde{U}$ as a subset of IB formed by the target source device. The mathematical description of the condition according to which block $\tilde{U}$ is issued by a specific source can be written as:

$$\exists \hat{u} \in U, |\hat{u}| = n, \ B(\hat{u}, S^{sec}) = 1,$$

$$\forall u \in U, \forall \hat{u} \neq u, |u| = |\hat{u}|, \ B(u, S^{sec}) = 0.$$
where:

\(- B -\) function for matching a subset of \( U \) to a value \( S^\text{key} \), if it is true this means there is an ordered set with power \( n \). In this case, the position of each \( R_i \) element of this \( S_i \) sets in the set itself is uniquely defined by the parameter \( S^\text{key} \) and the element content.

\(- n \) is the power of a subset \( U \), whose numerical value is uniquely determined by the information that the receiver has owned about the source.

\[ i = 1, \ldots, n \]

Figure 2. – Information blocks transmitted between the source and receiver.

When implementing the information source control system based on the IB index and hash function, the block allocation functions of the start and stop block are applied to blocks. Then the operation is applied to the remaining IB's selection information part, the operation of selecting the index \( f^\text{ind} \) and the operation of selecting the hash function \( f^\text{sh} \):

\[
\forall S \subseteq U \ s^\text{hash} = f^\text{sh} \left( S, S^\text{key} \right) S^\text{ind} = f^\text{ind} \left( S, S^\text{key} \right) \ 
\]

The recurrence rule for the formation of a subset that can be written in the form:

\[
R_0 = F^\text{sh} \left( s^\text{start} \right) = 0, \\
R_i = s^\text{ind} \implies s^\text{hash} = F^\text{hash} \left( R_{i-1} \right) \wedge R_i = s^\text{ind}, \\
R_{n+1} = n+1 \implies s^\text{stop} = F^\text{hash} \left( s^\text{stop} \right), 
\]

The membership function of a subset of the formation source will be written as:

\[
B \left( \tilde{u}, S^\text{key} \right) = 1 \iff \tilde{u} \left[ i \right] = s^\text{stop} \wedge j = n \wedge \tilde{u} \left[ j \right] = s^\text{stop} \wedge
\]

\[
\wedge f^\text{ind} \left( \tilde{u} \left[ i \right], S^\text{key} \right) = i, i = 1, \ldots, n \wedge \\
\wedge f^\text{sh} \left( \tilde{u} \left[ i \right], S^\text{key} \right) = F^\text{hash} \left( R^\text{sh} \left( \tilde{u} \left[ i \right], S^\text{key} \right) \right), i = 1, \ldots, n, u \subseteq U
\]

The implementation of the above method for forming a structured set of information blocks requires multiple hash comparison operations. The matches of the subset \( \tilde{u} \) blocks hashes with other hashes, leads to the forming tree structure of the blocks as shown.
in the Figure 3. This requires additional RAM costs for organizing the storage of the such structure. The purpose of this research is to obtain dependencies between the time and resource complexity of the implementation of the membership function $B$ over a random set of information blocks $U$ depending on the parameters of the algorithm for generating hash and index fields. The main parameters of the simulation:

- $M_{chain}$: The length of the legal chain block (The number of the authorized blocks).
- $I_{hash}$: The size of the hash function in bytes.
- $M_{false}$: The number of the unauthorized blocks.
- $W_{buff}$: The width of the buffer.

Figure 3. – A block tree that occurs as a result of matching the input block’s hashes with a block’s hashes in a source associated set.

4. THE RESULT OF THE STUDY

The result of a simulation study of a phased process of chain recovery is measurement the costs of RAM. It is obtained by conducting a series of repetitions in order to identify the percentage of detected errors. The experiment was carried out at 10,000 iterations. As a result of the study, we determine the dependencies of the size of the $I_{hash}$ and the number of $M_{false}$ and the $W_{buff}$.

The experiment was conducted with the $M_{chain}$ size 14, and information part size of a range of numbers from 0 to 1024. The experiment was conducted on 10,000 iterations with the following values: a $I_{hash}$ size is 5, 6 and 7 bits; the $W_{buff}$ is 14, 16 and 18; the number of $M_{false}$ is between (10-150) with an increment of 10 in each time. A graphical representation of the experiments per 10,000 iterations is shown in Fig. (4).
The result of the experiment is shown in Fig. (4). We see there is the relationship between the total number of chains and time. The volume of the total number of chains and time will increase with the increase of the number of $M_{false}$. On next stage we perform the same experiment, but we change the $W_{buff}$ to 16, 18 as shown in Fig. (5).

We note in this experiment as shown in Fig. (5), the volume of the total number of chains will be increased with the increasing of the number of $M_{false}$ and will be increased with the increasing of $W_{buff}$. If we perform the same experiment, but change the size of $I_{hash}$ to 6, we get the result shown in Fig. (6), and the result with $I_{hash}$ 7 as shown in Fig. (7).
Fig. 7. – The Study of the measure the cost of RAM ($M_{chain}=14$, $I_{hash}=7$, $W_{buf} (A=14; B=16; C=18)$).

As a result of the experiment as shown in Fig. (5-7), the total number of chains will increase with the increase of $M_{false}$ number. On the other hand, the total number of chains will be decreased with the growth of the $I_{hash}$.

Our simulation has shown that the cost of implementing a hash-based and index-based identification of an information block has a linear dependence on the parameter $M_{false}$ – numbers of elements of unauthorized information blocks set $U / \bar{u}$. This allows it to be used in the systems which use a high data exchange rates to form structured sequences of information blocks. The RAM memory costs are also a value that allows to use the method for the random length of $M_{chain}$ for a set of $\bar{u}$ blocks formed by a legal source.

5. CONCLUSION

As a result of the study of the change of the total number of buffered chains, no natural dependence was revealed. The empirical dependencies between buffer width and RAM costs are obtained. We can see, if the chain’s length does not change, even if the size of the hash changes, the total number of chains increases with the increasing of the unauthorized blocks number, and decreases with the increase of the hash size. Also, the study showed that the hardware cost for implementation the based on hash and index information block membership function has a linear dependence on the number of unauthorized blocks. This allows us to suggestion to be used in the systems which use a high data exchange rates to form structured sequences of information blocks.

References

[1] A. L. Maruhnenko, "Analysis of potential vulnerabilities and modern methods of protecting multi-user resources,” in infocommunication and space technologies: status, problems and solutions, Russia, Kursk, 2018.

[2] W. Liang, J. Xu, Y. Huang and S. Xie, "A distributed data secure transmission scheme in wireless sensor network,” International Journal of Distributed Sensor Networks , vol. 13, no. 4, pp. 1-11, 2017.

[3] M. A. Efremov, I. V. kaluutsky, M. O. Tanygin and I. I. Rudak, "Security of personal data, social networks and advertising in the global INTERNET,” News of Southwestern state University, Series: Management, computer engineering, computer science. Medical instrumentation., vol. 1, no. 22, pp. 27-33, 2017.

[4] A. L. Marukhlenko, M. O. Tanygin, M. A. Efremov and A. G. Spivakov., Security of information systems, Russia, Kurck, 2019, p. 210.

[5] O. R. M. Boudia and M. Feham, "Efficient Concealed Data Aggregation with Verification in Wireless Sensor Networks,” in IEEE International Conference on Advanced Networking Distributed Systems and Applications (INDS), 2014.

[6] C. Zhang, C. Li and J. Zhang, "A Secure Privacy-Preserving Data Aggregation Model in Wearable Wireless Sensor Networks,” Journal of Electrical and Computer Engineering, vol. 15, no. 1, pp. 1-9, 2015.

[7] W. Wang, D. Wang and Y. Jiang, Energy efficient distributed compressed data gathering for sensor networks, Ad Hoc Networks, 2016.

[8] V. Zaitsev and N. Sokolov, "Features of multiservice traffic taking into account messages generated by IoT devices,”
First mile, vol. 4, no. 1, pp. 44-47, 2017.

[9] O. N. Ledneva and E. P. Romasevich, “Analysis of the traffic of devices of the Internet of things,” Modern information technologies and IT education, vol. 14, no. 1, pp. 149-169, 2018.

[10] R. Sumathi, “A Secure Data Transfer Mechanism Using Single-Handed Re-Encryption Technique,” in International Conference on Emerging Trends in Science, Engineering and Technology, Tiruchirapalli, India, 2012.

[11] J. Yin, Y. Yang, L. Wang and X. Yan, “A reliable data transmission scheme based on compressed sensing and network coding for multi-hop-relay wireless sensor networks,” Computers and Electrical Engineering, 2016.

[12] P. Hao and Q. Shi, “Matrix factorizations for reversible integer mapping,” IEEE Transactions on Signal Processing, vol. 49, no. 10, pp. 2314 - 2324, 2001.

[13] P. S. I. Group, “PCI Express® Base Specification Revision 3.0,” [Online]. Available: URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.694.7081&rep=rep1&type=pdf. [Accessed 2019 05 15].

[14] M. O. Tanygin and A. P. Tipikin, “Methods of authentication of information protection systems and controlling software,” Telecommunications and Radio Engineering, vol. 66, no. 5, pp. 453-463, 2007.

[15] M. A. Ivanov, Cryptographic methods of security information in computer systems and networks, Moscow: MEPhI, 2012, p. 400.

[16] S. Azam, R. Manzoor and M. u. Rehman, "Secure Solution to Data Transfer from Sensor Node to Sink against Aggregator Compromises," in 9th International Conference on Frontiers of Information Technology FIT 2011, Islamabad, Pakistan, 2011.

[17] M. Bellare, J. Kiliany and P. Rogawayz, "The Security of the Cipher Block Chaining Message Authentication Code," Journal of Computer and System Sciences, vol. 61, no. 3, p. 362–399, 2000.

[18] M. O. Tanygin and A. P. Tipikin, "System architecture for hardware restriction of access to information on a computer hard disk," Telecommunications, vol. 3, pp. 44 - 46, 2006.

[19] M. O. Tanygin, H. Y. Alshaeaa and V. Altukhova, "Establishing Trusted Channel for Data Exchange between Source and Receiver by Modified One-time Password Method," in International Russian Automation Conference, RusAutoCon, Sochi, Russian Federation, 2019.

[20] J. Black, Authenticated Encryption. In Encyclopedia of Cryptography and Security, Tilborg: Springer, 2005.

[21] V. P. Dobrisa, A. Plugaru, A. A. Zakharunya and L. A. Eivanova, "An Approach to Creating an Adaptive Algorithm for Transmitting Information in Open Channel with Mutual Authentication of the Source and Receiver," in Intelligent Information Systems: Trends, Problems, Prospects; Materials of Reports of the VI all-Russian Scientific and Practical Conference "IIS-2018"., Kursk, 2018.

[22] M. O. Tanygin, M. A. Efremov and H. Y. Alshaeaa, "Analysis of the Secure Data Transmission System Parameters," in International Russian Automation Conference, RusAutoCon, Sochi, Russian Federation, 2019.

[23] M. Tanygin, "Method of Control of Data Transmitted Between Software and Hardware," in Computer Science and Engineering: Materials of the IV International Conference of Young Scientists CSE-2010, Lviv, 2010.

[24] M. O. Tanygin, H. Y. Alshaeaa and V. A. Altukhova, "A Method of Control the Integrity of Transmitted Blocks Information," Telecommunications, vol. 3, pp. 12-21, 2019.