UDC 004.052.42

ZERO-KNOWLEDGE IDENTIFICATION OF REMOTE USERS BY UTILIZATION OF PSEUDORANDOM SEQUENCES

I. Daiko, V. Selivanov, M. Chernyshevych, O. Markovskyi

The article theoretically substantiates, proposes and investigates an identification scheme based on the concept of "zero knowledge" using irreversible generators of pseudorandom bit sequences. Session passwords form a chain generated by selective sequence values. Secondary identification sessions are provided in the proposed scheme to counter attacks with the displacement of one of the remote interaction parties. The main elements of the proposed identification scheme are developed in detail: authorization procedures, primary and secondary identification.

Key words*: Zero-knowledge identification, chain of passwords, cryptographically strong identification, generators of pseudo-random bit sequences, middle attacks.*

Introduction

The high rate of progress in the technical means of the Internet, as well as the COVID-19 pandemic, have resulted in the rapid spread of remote interaction technologies. On the other hand, progress in the field of nanotechnology has stimulated the dynamic expansion of the use of remote computer control systems for real-world objects. A characteristic feature of such systems, which have received the name Internet of Things (IoT), is that the Internet is used as a data transmission medium [1]. The above-mentioned expansion of the use of remote information interaction systems to new areas of human activity stimulates the corresponding growth of computer crimes, the purpose of which is to affect the processes of data exchange between the parties of such interaction [2].

This requires adequate improvement and development of means of protection of processes of remote information interaction. Mechanisms for mutual identification of remote interaction participants occupy an important place among these tools. In recent years, there has been a significant qualitative and quantitative increase in attacks on these mechanisms, and new forms of violation of identification processes emerged. Accordingly, there is an objective need for the improvement of means of identification of the parties of remote interaction to ensure the proper level of data protection, differentiation of access rights to them, and the efficiency of implementing economic forms of the organization of remote provision of information services.

The problem of reliable identification is particularly acute for systems of computer remote control of objects in the real world using the Internet as a data transmission medium. The terminal devices of such remote-control systems are portable microcontrollers with built-in radio modems. These computing devices have low computing power, but must implement real-time identification procedures. New methods of fast and reliable identification need to be developed for them.

Thus, the scientific task of improving the effectiveness of means of identification of participants in remote information interaction is relevant and practically crucial given the peculiarities of the current stage of information technology development.

Problem statement and review of methods for its solution

The effectiveness of identification mechanisms, like any other means of cryptographic data protection, is characterized by two criteria [3]:

- the level of security, which is estimated by the number of resources needed to breach protection;
- the amount of resources for the implementation of protection functions. As the last criterion, the time of execution of protection functions on the computing platform of the participant of information interaction is most often used.

Traditionally, the task of identifying a remote participant in information interaction is one of the three fundamental tasks of modern cryptography [4]. The analysis of this problem is based on the classic model of remote information interaction. This model takes the presence of a system that remotely provides definite information services to a certain number of subscribers. This means that

the specified classical model assumes an asymmetric nature of threats: the motivation to obtain illegal access to system resources is much higher than the motivation to provide services to the user instead of the system. Accordingly, within the framework of the classic model, the identification mechanisms are also asymmetric in terms of both the level of security and the speed of implementation of protective functions: for a system that serves thousands of subscribers, the identification time should be orders of magnitude longer. Such a model sufficiently correctly reflects a wide range of real multi-user systems, as well as a significant nature of threats to the information security of computer management systems of real-world objects [5]. An essential element of the model is that data exchange is carried out over potentially vulnerable Internet data transmission channels.

Within the classical model discussed above, the goal of attacks on identification mechanisms is to gain illegal access to system resources, to obtain information exchanged between the system and users, or to change it intentionally. The objects of the attack are the data exchange channel between the user and the system and, in particular, Internet switching centers [6].

A passive attack on the channel involves control of the data transmitted over the channel during the identification process. Active action on the channel involves interception of the data exchange session after the system identifies the user (middle attacks).

Another object of attack is the system in which the identification data of its users is stored. Technologically, the attack on the system is mainly carried out under the guise of a legal user, virus programs, or mafia fraud, that is, by influencing the system personnel [7].

All known methods of identification of remote participants of information interaction are divided into two classes: cryptographically strict and cryptographically weak [3].

Cryptographically strict identification methods must satisfy the following conditions:

- The password must be changed in each session of information interaction in order to protect against passive attacks on the channel and attacks on the system in which passwords are stored.
- The system should not store any information that allows for the reproduction of subscribers' passwords.

Accordingly, weak identification methods use permanent passwords that can potentially be intercepted during transmission in the channel and used to illegally penetrate the system under the guise of a legitimate user. A class of hybrid identification methods can be singled out separately, which fulfill only one of the above conditions. This class includes, in particular, the mechanism for identifying users of UNIX systems [3], within which only the second condition is satisfied.

In practice, cryptographically strict identification is implemented most often in the form of the concept of "zero knowledge" [8], which is based on two provisions:

- the subscriber must have a cryptographic mechanism for generating correct passwords;
- the system must have at its disposal a cryptographic mechanism for checking the correctness of passwords, which, however, does not allow the system itself to generate correct passwords.

Until now, a wide range of means of identification has been proposed within the framework of the concept of "zero knowledge" [9-12], which uses various cryptographic mechanisms for the formation of correct passwords and their verification by the system.

Existing methods of cryptographically strict identification can be divided into two classes:

- with unrelated session passwords;
- with session passwords related to each other.

In the identification schemes of the first type, mathematical multi-valued irreversible transformations are used as a mechanism for checking the correctness of the subscriber's password. In the well-known Guillou-Quisquater [9], Schnorr[10], and FESIS [11] schemes, irreversible number theory transformations are used as such transformations. In these schemes, the impossibility for the system to independently generate the correct session passwords is due to the analytical intractability of the discrete logarithm problem. The possibility of using a large number of independent session passwords is because this problem has an infinite set of solutions [3].

On the other hand, the use of number theory problems to build a mechanism for verifying the correctness of a password has the consequence of spending considerable time on their implementation, given the high computational complexity of performing modular exponentiation of large-bit numbers.

A particular increase in identification speed while preserving the above-mentioned cryptographic properties can be achieved using the algebra of finite Galois fields [12], especially in hardware implementation.

Significantly greater opportunities for fast cryptographically strict identification are provided by schemes with associated session keys. In fact, if when using independent session passwords, the user proves that he is the same one who registered in the system, then with dependent passwords, he demonstrates that he is the same one who interacted with the system in the previous information interaction session.

A classic identification scheme of this type [13] involves the use of an irreversible hash transformation. At the registration stage, the subscriber forms a chain of session passwords, each resulting from a hash transformation over the previous one. Accordingly, the subscriber uses these passwords in reverse order, so the system has the last password session as an identification code. Due to the irreversibility of the hash conversion, it cannot determine the next session password. Using such a scheme provides 3-4 orders of magnitude faster identification than the methods discussed above based on irreversible transformations of number theory.

An analysis of the current practice of attacks on remote interaction systems shows that when using cryptographically strict identification schemes, the biggest threat is middle attacks [14]. These attacks are carried out after the subscriber's identification is completed and consist in pushing him away from informational interaction with the system.

The most effective way to counter these types of attacks is to carry out repeated identification cycles during the information interaction session. The review showed that the most significant disadvantage of known cryptographically strict identification schemes in current conditions is vulnerability to middle attacks.

Purpose and objectives of research

The purpose of the work is to increase the effectiveness of cryptographically strict identification of participants in remote information interaction due to the acceleration of the identity confirmation process, as well as by organizing secondary cycles of contact control to counteract interaction interception.

To achieve the set goal, the following tasks are solved in the work:

- analysis of the possibilities of use for cryptographically strict identification of the fastest-acting standardized cryptographic mechanisms – generators of pseudo-random binary sequences;
- development and research of a method of cryptographically strict identification, which is distinguished by the use as a mechanism for checking the correctness of the session password on the side of the system of properties of irreversible generators of pseudorandom binary sequences, due to which, an increase in speed is achieved and the possibility of implementing a series of secondary accelerated identification cycles using them;
- theoretical and experimental evaluation of the effectiveness of using generators of pseudo-random binary sequences as a mechanism for checking the correctness of the session password in terms of speeding up the identification process, as well as increasing the level of security.

The object of research is the process of cryptographically strict identification of participants in remote information interaction, which provides the possibility of protection against session interception by outsiders.

The method of implementing the concept of "zero knowledge" using pseudo-random sequences for subscriber identification

In current conditions and in the future, the level of security of identification processes acceptable for most applied applications can be achieved only by applying the progressive cryptographic concept of "zero knowledge." However, the main problem with the practical use of these technologies is the need for significant dusting resources to implement the corresponding cryptographic transformations.

A high speed of cryptographically strict identification can be achieved only when nonlinear Boolean transformations are used as irreversible transformations. It is well known that the system of nonlinear Boolean equations cannot be solved by analytical methods [3]. The only way to solve such systems of nonlinear Boolean equations is to perform a complete enumeration.

Cryptographic transformations implementing the concept of "zero knowledge" using non-linear Boolean functions can be carried out by three known mechanisms:

- one-way hash transformations;
- cipher blocks that are used in the mode of one-way transformations;
- generators of binary pseudorandom sequences used in stream ciphers.

It is well known that current ciphers [4] provide the highest speed of cryptographic protection of information. They are widely used for real-time encryption and decryption of telephone conversations and transmission of video images over closed channels. The main advantage of using pseudo-random sequences as irreversible transformations for implementing the cryptographic concept of "zero knowledge" is a significantly faster performance than hash transformations and cipher blocks.

The developed method of rapid identification of remote interaction participants involves using a generator of pseudorandom bit sequences by the system and each remote subscriber. Cryptography uses generators that remember the state, non-linear Boolean functional transformations of the transition to the next state, and the formation of the output bit. In other words, the generator circuit fits into the well-known abstract automaton model [15]. The fundamental point here is that the Boolean functional transformations used have high nonlinearity and meet the criterion of the avalanche effect. This makes it impossible to reconstruct the sequence of bits of the line using the methods of linear and differential cryptanalysis [8]. Unlike the traditional one, the automaton model of the generator of pseudorandom bit sequences has no input signals; that is, the mathematical model of the generator is an abstract Moore automaton. This means that the line of following the generator states is determined uniquely with fixed settings of the feedback functions. At the same time, the feedback functions are organized so that the line of states has a maximum period of τ and includes all possible conditions. This means the generator of pseudorandom binary sequence forming a bit sequence of $B = b_1, b_2, \ldots b_\tau, b_1, b_2, \ldots$ with a repetition period τ.

The use of pseudorandom binary sequences in modern cryptographic data protection mechanisms is based on the practical difficulty of restoring the entire series by its k-bit fragment b_1, b_2, …, b_k. The complexity of this problem is due to the nonlinearity of the output bit formation function based on the status code of the generator, which transforms the task of restoring the sequence into a system of nonlinear Boolean equations that cannot be solved analytically. This means that the only way to restore the series by its fragment is an enumeration, which with current values of the period τ goes far beyond the possibilities of technical implementation. In a practical sense, this means that knowing the algorithm of the pseudorandom bit sequence generator for its given fragment, it is impossible to restore the state of the generator, starting from which the given segment is generated.

The developed method of cryptographically strict identification of participants of remote information interaction involves the procedures of system subscriber registration, identification at the beginning of the session, and secondary identification, which is carried out periodically during the current session.

The first of the mentioned procedures is that a chain of m session passwords is formed, stored in the subscriber's memory, and its first element is sent to the system that performs identification. The procedure is performed in the following order:

1. The subscriber randomly generates the code S_m of the generator status of pseudo-random sequences on the m-th identification session. The setting of the Boolean function of the output signal's formation is performed, as well as the setting of the feedback functions of the generator.

2. The selected code S_m is loaded into the generator's memory, after which a sequence of k bits is formed, which form the session password S_{m-1}.

3. The value of m decreases by one: $m = m - 1$. If after that $m > 0$, a return is made to the execution of the previous item 2.

4. The pseudo-random bit sequence generator setting codes chosen by the subscriber and the S_0 code are encrypted with the system's public key and sent to it. The generated sequence of codes S_1, S_2, …, S_m of the generator state is stored in the subscriber's memory.

5. The system receives the registration message from the subscriber, decrypts it with its private key, and stores the value of the codes of the subscriber's generator settings and the S_0 code in the memory area allocated for the subscriber's service.

The developed subscriber identification procedure at the beginning of the i-th session of information interaction, $i \in \{1, 2, \ldots, m\}$ includes the following actions:

1. The subscriber sends the i-th S_i code to the system, which plays the role of a session password. In addition, the subscriber initiates the operation of the generator of pseudo-random bit sequences with the start code S_i and forms a line with a length of $k + d$ bits. The first k bits of the generated series constitute the X code, and the last d bits set the U code.

2. The remote system receives an initialization code from the subscriber via the Internet, by which it selects the setting codes for the generator to work with the subscriber from memory. Then the system adjusts the generator by the read codes. After receiving the session password code Si from the subscriber, the system initiates the operation of the generator of pseudo-random bit sequences with the start code S_i and forms a line of length $k + d$ bits. The first k bits of the formed series form the Y code, and the last d bits form the W code.

3. The system compares the Y code generated from the generator operation with the S_{i-1} code of the previous session password stored in the memory. Suppose these codes are identical, i.e., $S_{i-1} = Y$. In that case, the primary identification of the subscriber in the current session is considered successful, and system resources determined by his status are provided to him. The S_{i-1} code in the system memory is replaced by the accepted session password S_i. In addition, the system sends the user the code W generated by it. If $S_{i-1} \neq Y$, the system sends a zero code to the subscriber.

4. The subscriber receives the access rights granting code from the system: if it is zero, the system has not identified it. Otherwise, the received code W is compared by the subscriber with the code U: if $U = W$, then the user gets confirmation that he has a remote interaction session with the system. In this case, he starts an information interaction session.

The developed procedure for the basic identification of the subscriber by the system corresponds to the above criteria of the theoretical concept of "zero" knowledge. The first of these criteria is satisfied because the session password codes $S_1, S_2, \ldots, S_m$ used by the subscriber are all different. In practice, this is guaranteed by the appropriate choice of their length - k. It is also quite evident that the system, having the S_{i-1} code at its disposal, is not able to obtain the code of the next session password, even knowing the settings of the Boolean functions of the output signal formation and the feedback function of the generator of pseudo-random bit sequences. It is not difficult to show that the specified problem in the mathematical sense is identical to the solution of a system of k nonlinear Boolean equations, which cannot be solved analytically. Thus, the second criterion of the concept of "zero" knowledge is fulfilled: the system cannot generate the subscriber's correct session password.

To protect against attempts to intercept the process of information interaction between the system and the subscriber after its identification by blocking at the switching centers of global networks, the procedure for intermediate identification of the participants of the information interaction has also been developed within the framework of the proposed method, which makes it possible to periodically check the identity of the participants of the information interaction. This procedure consists of the following sequence of actions:

1. The system periodically, without additional generator setting, forms a $2 \cdot d$-bit pseudo-random sequence, the first d bits of which form the Q code and the last d - the G code. The system sends the G code obtained in this way via the Internet to the subscriber with whom the information interaction is carried out.

2. The subscriber receives the G code from the system and initiates re-identification. For this, the subscriber, without additional adjustment of his generator, generates a $2 \cdot d$-bit pseudo-random sequence, the first d bits of which form the V code, and the last d bits constitute the C code. After that, the subscriber compares the generated code C, and the code G received from the system: if $G = C$, then this means that the system supports information interaction with it. In this case, the subscriber sends the V code he generated to the system.

3. The system receives the V code from the subscriber and compares it with the Q code generated using the system generator. If these codes match, i.e., $V = Q$, then the system makes sure that the information interaction takes place in the subscriber and that he was not pushed out of the session by an intruder.

Thus, the use of the described procedure for secondary identification of participants in remote information interaction allows you to reliably detect the presence of attacks on a session after its start. It is quite obvious that the resource costs for secondary identification are an order of magnitude lower compared to primary identification. This allows for secondary identification during the session without appreciable impact on the speed of data transfer between participants.

Effectiveness evaluation of the method

The main criteria for the effectiveness of systems for identifying participants in remote information interaction are the level of security and the speed of technical implementation of protection procedures.

The task of breaking the proposed method of cryptographically strict identification is identical to the task of predicting a pseudorandom binary sequence. When using standardized cryptographic generators of pseudo-random bit sequences, this task is one whose practical implementation is beyond technical capabilities [4]. In order to break the security of the system, that is, to simulate access to it by a legitimate user, it needs to determine two components: the number h_x of sequence generation steps and the initial state Q_x. This can only be done by sorting.

The possibility of implementing basic and repeated identification cycles using a single mechanism - a cryptographic generator of pseudorandom bit sequences- is the main advantage of the proposed method of cryptographically strict identification. Reputed cryptographically strict identification schemes [9-12] do not provide such a possibility. Another significant practical benefit of the developed technique of cryptographically rigorous identification compared to noted ones is much faster performance.

In the software implementation of the SHA-256 standardized hash transformation, the most common in practice, 64 cycles are performed, in each of which 6 shifts, 15 arithmetic addition operations, 5 logical AND operations, 6 logical XOR operations are performed, i.e. a total of 2048 operations.

With the software implementation of the AES cipher block in the minimum configuration, 10 cycles are performed, in each of which the logical XOR operation of the data block with the key is performed (4 processor operations), 16 operations of accessing the byte substitution tables of the data matrix, 3 cyclic shift operations for shuffling the rows of this matrix, 8 shift and logical addition operations to shuffle the columns of the data matrix. The total number of processor operations is therefore 310 processor operations.

Structures of generators of pseudorandom generators are much simpler and dozens of processor operations are used to generate one bit, which is much less compared to cipher blocks or hash transformations.

It is proved [3] that block ciphers and hash converters, which are used in known cryptographically strict identification schemes, have a lower speed order of magnitude than generators of pseudorandom bit sequences.

Unlike known methods, only one pass of a password is used for a single identification session over potentially dangerous data lines.

The proposed method of cryptographically strict identification allows to detect the fact of an attempt to oust a legal participant of information interaction from the session. The attacker does not know the settings of the pseudorandom sequence generator, so he cannot generate the correct sequence of bits that confirms the presence of an information contact. However, the developed method does not protect the process of remote interaction for types of attacks in which the attacker fully controls the information flows between the interaction parties at the switching center. To implement such protection, it is necessary to use a generator of pseudo-random sequences for stream encryption of data exchanged by participants of information interaction. The use of a generator of pseudorandom binary sequences within the framework of the proposed solution allows the use of a single cryptographic mechanism for data identification and encryption.

Conclusion

Conducted research aimed at increasing the effectiveness of cryptographically strict identification of participants in remote information interaction allowed to obtain the following results:

It has been established that the main shortcomings of existing schemes of cryptographically strict identification in modern conditions are insufficient speed, as well as the inability to resist new types of attacks, in particular, displacement of the subscriber from the process of remote information interaction after his identification by the system.

A method of cryptographically strict identification has been developed and researched, which is distinguished by the use as a mechanism for checking the correctness of the session password on the system side of the properties of irreversible generators of pseudorandom binary sequences, due to which we achieve an increase in performance and the possibility of implementing a series of secondary accelerated identification cycles using them;

The proposed method of accelerated identification of participants of remote information interaction is oriented for use in real-time computer control systems of remote objects using global networks.

References

1. Noot M.M. Current research on Internet of Things (IoT) / M.M. Noot, W.H. Hassan // Compute Network. -Vol.148. -No.15. -2019. -pp.283-294.
2. Khan M.A. IOT security: Review, blockchain solution and open challenges / M.A. Khan, K. Salah // Future Generation Computer Systems. -Vol. 82. -No.5. -2018. -pp.395-411.
3. Schneier B. Applied Cryptography. Protocols, Algorithms and Source Code in C / B. Schneier // Wiley. -2015. -P.784.
4. Menezes A.J. Handbook of Applied Cryptography / A.J. Menezer, P.C. Van Oorschot, S.A. Vanstone // CRC-Press. – 1997. – 780 c.
5. Mashal I. Analysis of recommendation algorithms for Internet of Things / I. Mashal, O. Alsaryrah, T.Y. Chung // IEEE Wireless Communications and Networking Conference Workshops (WCNCW), -2016. -pp. 181–186.
6. Kittichokechai K. Secret Key-based Identification and Authentication with a Privacy Constraint / K. Kittichokechai, G. Caire // IEEE Trans. Inf. Theory. -Vol. 62. - 2016. - № 11. -P. 6189-6203.
7. Cheng P. Zero-knowledge identity authentication for internet of vehicles: Improvement and application / M. Han, Z. Yin, P. Cheng, X. Zhang, S. Ma // PLoS ONE. -2020. - Vol.15 – No.9. -P.217-247.
8. Bardis N. Fast subscriber identification based on the zero knowledge principle for multimedia content distribution / N. Bardis, N. Doukas, O.P. Markovskyi // International Journal of Multimedia Intelligence and Security. - 2010. - No.4,- P. 363-377.
9. Quisquater J.J. A Practical Zero-Knowledge Protocol Fitted to Security Microprocessor Minimizing Both Transmission and Memore. Proceeding of Advances of Cryptology / L.C. Guillou, J.J. Quisquater // Eurocrypt -88.- Springer - Verlag. - 1988.- P.123-128.
10. Schnorr C.P. Method for Identification Subscribers and for Generating and Verification Electronic Signatures in data Exchange System / C.P. Schnorr // US Patent #4995,083.19-1991.
11. Feige U., Fiat A., Shamir A. Zero knowledge proofs of identity / U. Feige, A. Fiat, A. Shamir // Journal of Cryptology, - 1988.- Vol.1.- №.2. – P.77-94.
12. Bardis N.G. A Method for strict remote user authentication using non-reversible Galois field transformations / N.G. Bardis, N. Doukas // MATEC Web of Conferences. Vol. 125, - 2017.- pp.243-249.
13. Asimi Y. Strong zero-knowledge authentication based on the session keys (SASK) / Y. Asimi, A. Amghar, A. Asimi, Y. Sadgi // International Journal of Network Security & Its Applications (IJNSA). -2015.- Vol.7, - No.1, - pp.51-66.
14. Conti M. A Survey of Man in the Middle Attacks. / M. Conti., N. Dragoni, V. Lesyk //. IEEE Communications Surveys and Tutorials.- 2016. -Vol.18.- № 3.- pp.2027-2051.
15. Unkaševi´ T. A Generic Model of the Pseudo-Random Generator Based on Permutations Suitable for Security Solutions in Computationally-Constrained Environments./ T. Unkaševi´, Z.Banjac, M. Milosavljevi´c // Sensors 2019,- Vol.19, - P.5322-534