Modeling of the User’s Identification Security System of on the 2FA Base

Olga Ussatova, Saule Nyssanbayeva, and Waldemar Wójcik

Abstract—The article describes methods of user identification using authentication based on the second factor. Known algorithms and protocols for two-factor authentication are considered. An algorithm is proposed using mobile devices as identifiers and generating a temporary password based on the hash function of encryption standards. For an automated control system, a two-factor authentication model and a sequential algorithm for generating a temporary password using functions have been developed. The implementation of the system is based on the Node.js software platform using the JavaScript programming language, as well as frameworks and connected system libraries. MongoDB, an open source database management system for information storage and processing was used.

Keywords—two-factor authentication, data security, user identification, password generation

I. INTRODUCTION

The dynamics of the development of information technology in the socio-economic and cultural life of society and the state places high demands on the solution of information security issues.

Therefore, since the independence of the Republic of Kazakhstan, the first president, Nursultan Abishevich Nazarbayev, has repeatedly focused on the need to protect the interests of the people living in the republic. Thus, in the Address of the President of the country to the people of Kazakhstan dated October 10, 1997, “Kazakhstan is 2030. Prosperity, Security, and Improving the Well-Being of All Kazakhstans,” national security is defined as a long-term priority, one of which is information security [1].

Ensuring the information security of the state, in turn, requires the use of an integrated approach, including organizational, technical, software, social mechanisms that can realize the constitutional rights and freedoms of man and citizen in the field of obtaining information, using it in order to protect the constitutional system, sovereignty and territorial integrity of the Republic of Kazakhstan political, economic and social stability, law and order, development of mutually beneficial international cooperation in the field of information security

With the development of information and telecommunication technologies, our society has acquired a huge amount of benefits. There was an opportunity to make remote purchases of goods from anywhere in the country, to monitor the status and execution of work, without being present at the places of their conduct. Currently, it is possible to issue almost any certificate without leaving home, thanks to an electronic digital signature and the information portal of public services. But, unfortunately, with the development of technological progress and the general level of informatization, new threats appear. Cybercrime allows cybercriminals to commit unlawful and illegal actions, being thousands of kilometers from the target of their attack.

In the Message to the people of Kazakhstan “Third Modernization of Kazakhstan: Global Competitiveness”, the President of the Republic of Kazakhstan noted that the fight against cybercrime is becoming increasingly relevant [2].

In the light of the implementation of the above Message, taking into account the approaches of the Kazakhstan-2050 Strategy for Kazakhstan becoming one of the 30 most developed countries in the world, the Concept of Cybersecurity (“Cyber Defense of Kazakhstan”, dated June 30, 2017, was developed and approved by the Government of the Republic of Kazakhstan), where the main directions of the implementation of state policy in the field of protection of electronic information resources, information systems and telecommunication networks, ensuring the safe use of information technologies.

The development of information technology is gaining momentum every year. Information technology has become an integral part of our life, in connection with this there is an automation of all processes of human life. Most of the information is automated and stored in information systems, which must be protected.

Attacks on information have become commonplace. Attackers use both errors in writing and administering programs, and methods of social psychology to obtain the desired information. Resource developers who are supposed to work with user data are required to protect this data and prevent the possibility of leakage.

One of the most effective methods of information security today is two-factor authorization for entering the system. It involves double data protection by linking the account to the protection system. After binding, the user will need to interact with this system to verify the data.

Identification and authentication of subjects are the main means of protecting an information object from outside interference. The verification procedure can be carried out both at the entrance to the information system, and inside it, when moving to a new data object. Modern means of protection against unauthorized access implement complex algorithms for

O. Ussatova is with Al-Farabi Kazakh National University, Almaty, Kazakhstan and Institute of Information and Computational Technologies, Almaty, Kazakhstan (e-mail: uoa_olga@mail.ru).

S. Nyssanbayeva is with Institute of Information and Computational Technologies, Almaty, Kazakhstan (e-mail: sultashai@mail.ru).

W. Wójcik is with Lublin University of Technology, Nadbystrzycka 38a, 20-618 Lublin (e-mail: Waldemar.Wojcik@pollub.pl).

© The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.
the analysis of the distinctive features of the subject and his behavior. To strengthen protection against external threats, strict verification rules are established or additional measures are applied to double-check the results. This security policy enhances the reliability of the security system.

II. GOALS AND OBJECTIVES OF THE STUDY

The purpose of the work described in the article is the development and study of information protection procedures to ensure the integrity of electronic documents during their storage and exchange in databases of automated systems (AS) using user identification based on the second factor.

To achieve this goal, the following tasks were set and solved:
1. The study of existing algorithms and protocols for multifactor authentication.
2. The use of databases and methods for their protection.
3. Development of a model for protecting information in databases of automated systems based on two-factor authentication.

The object of research is a subsystem for protecting information from unauthorized access when identifying a user based on the second factor.

The novelty of the proposed system:
- generation of a set of functions for obtaining the result for each individual system;
- development of a modified algorithm for generating a secret one-time code;
- introduction into a system closed from the outside world;
- scalable complexity of functions for computing a one-time key.

III. SECURITY METHODS USING TWO-FACTOR AUTHENTICATION

Modern speakers have disadvantages associated with problems of integrity, confidentiality and accessibility. Information protection (ZI) in an automated control system is, first and foremost, the right approach to system design, by which you need to understand the whole variety of process control systems. In this case, it is necessary to take into account all the regulatory requirements for information security and use data transfer protocols allowing various options for all kinds of attacks and methods of protection against them. After research and systematization of knowledge, a threat model and a model of system violators are compiled, which allows you to choose the most appropriate protection methods. Moreover, when using the correct approach to the design and application of a set of tools that are reliably integrated with each other, the ZI system will not be visible to the user, and its functioning will not affect the speed of the entire system.

All the various means of information protection should be included in any industrial system as one of the necessary subsystems - the subsystem ZI. When implementing the subsystem, security gateways, integrated software or hardware tools for cryptographic information protection, attack detection systems, and other ZI tools can be used. Almost all automated control systems (ACS) use a role-based access system to one or another object management function, and operator actions are usually logged in the corresponding journal. When two additional elements are introduced into this function - two-factor authentication and the use of digital cryptographic signatures - the data can be not just data, but a legally important document. Then, when clarifying the circumstances of a particular case, it will be known for certain who became the source of the unbalancing control effect, which led to the subsequent negative consequences.

Two-factor authentication is a security algorithm in which the user offers two different authentication factors, which improves access protection to both user credentials and user resources [1]. Traditional systems use a username and password for authentication. This method provides a minimum level of security, as names and passwords can be easily intercepted and even guessed. Two-factor authentication provides a higher degree of protection than one-factor authentication, in which the user offers only one factor, usually a password, and is used to control access to sensitive systems and data. Consider some well-known protocols and authorization algorithms:

OAuth- an authorization protocol that allows you to grant a specific service or application access rights to user resources on a third-party service or application [2]. This protocol provides an opportunity to get rid of the need to trust a third-party application for authorization data. This protocol allows you to issue some specific set of rights, and not all at once.

The implementation of the OAuth protocol is based on the use of basic web technologies such as HTTP requests and redirects (route redirection). As a result, its use is possible on various platforms with access to the Internet and a browser. The general structure of the application using the OAuth protocol includes two stages:
- obtaining authorization;
- access to protected resources.

The result of authorization is an “access token” - a specific key, the presentation of which opens access to protected resources. In the simplest case, they are accessed via HTTPS with the received access token indicated in the headers or as one of the parameters.

The protocol describes several authorization options for various situations:
- authorization for applications that have a server part, most often these are sites and web applications;
- authorization for fully client applications;
- authorization by login and password;
- restoration of the previous authorization.

The general principle of working with the OAuth protocol is as follows:
- redirection of the authorization page;
- on the authorization page, the user is requested confirmation of the issuance of rights;
- in case of user consent, the browser redirects to the URL specified when opening the authorization page, with the addition of a special key in the GET parameters - the "authorization code";
- the application server performs a POST request with the obtained authorization code as a parameter, as a result of which the access token is returned.

OAuth security is largely based on SSL. This greatly simplifies the life of developers, but requires additional computing resources and administration. This can be a significant issue in highly loaded projects.
HOTP (HMAC-Based One-Time Password Algorithm) – a secure authentication algorithm using a one-time code based on SHA-1. This algorithm is a one-way authentication algorithm, which involves server-side client authentication [3]. The algorithm was first formally described by the IETF team in December 2005. The parameter responsible for the dynamics of password generation is the fact of generation or the event itself. Each time a new password is created, the event counter increments by one. It is such a monotonous increasing value that is used as the main parameter of the algorithm. The second parameter responsible for generating one-time passwords is a symmetric key, which must be unique for each client and at the same time closed to everyone except the server and the client itself. Protection systems built using HOTP are highly reliable. They are mainly resistant to widespread cryptographic attacks.

The HOTP algorithm is based on increasing the value of the counter of a static symmetric key, known only to the token and validating the provision of services. To create a HOTP value, the HMAC Algorithm SHA-1 is used, as defined in RFC 2104 [3].

Since the output of the HMAC-SHA-1 calculation is 160 bits, you need to reduce (truncate) the value entered by the user.

$$\text{HOTP}(K, C) = \text{Truncate}(\text{HMAC-SHA-1}(K, C))$$  \hspace{1cm} (1)

where Truncate is a function that converts an HMAC-SHA-1 value to a HOTP value, $K$ – key; $C$ – counter

The values generated by the HOTP generator are considered large inverse byte order. HOTP value generation is described in 3 stages:

1. Generate HMAC-SHA-1 value, let

$$\text{HS} = \text{HMAC-SHA-1}(K, C),$$  \hspace{1cm} (2)

2. Generating a 4-byte string (dynamic truncation), let

$$\text{Sbits} = \text{DT}(\text{HS}),$$  \hspace{1cm} (3)

where: $\text{DT}$ returns a 31-bit string.

3. Calculate the HOTP value, let

$$\text{Snum} = \text{StoNum}(\text{Sbits}),$$  \hspace{1cm} (4)

where: $S$ - is converted to a number $v_0 \ldots 2^{31} - 1$.

$$\text{Return } D = \text{Snum} \text{ mod 10Digit},$$  \hspace{1cm} (5)

where: $D$ is the number in the range $0 \ldots 10^{\text{discharge}} - 1$.

The truncation function performs stages 2 and 3, that is, dynamic truncation and modulo 10 reduction. The purpose of the method is to extract a 4-byte dynamic binary code from the 160-bit (20-byte) result of HMAC-SHA-1.

The reason for masking the most significant bit of $P$ is to avoid confusion in modulo calculations. Different processors perform these operations in different ways, and masking the signed bit removes all the uncertainty. The minimum value is a 6-digit code.

TOTP (Time-based One-Time Password Algorithm) - an algorithm for creating one-time passwords for secure authentication (2008) [4]. It, like HOTP, is a one-way authentication algorithm in which the server authenticates the client. In contrast to HOTP, the non-parameter responsible for the dynamics of password generation is time. Usually not a specific time indication is used, but the current interval with predetermined boundaries. The concept of one-time passwords, coupled with modern cryptographic methods, can be used to implement reliable remote authentication systems. TOTP is quite resistant to cryptographic attacks, but there is a chance of hacking. For example, such an option is possible as a “man in the middle”.

A TOTP implementation may use the HMAC-SHA-256 or HMAC-SHA-512 functions based on the SHA-256 or SHA-512 hash functions.

The TOTP algorithm is a variant of the HOTP algorithm based on the representation of the counter as a time factor.

$$\text{TOTP} = \text{HOTP}(K, T),$$  \hspace{1cm} (6)

where: $T$ is an integer and represents the number of time steps between the initial counter, the time $T_0$ and the current time of the operating system (OS). Then

$$T = (\text{Current time OS } T0)/X,$$  \hspace{1cm} (7)

where: floor function is used by default in calculations.

For example, with $T_0 = 0$ and a time step of $X = 30$, $T = 1$ if the Current time OS is 59 seconds, and $T = 2$ if the Current time OS is 60 seconds. The implementation of this algorithm should support a time value $T > 32$-bit integer when it goes beyond 2038. The values of the system parameters $X$ and $T_0$ are pre-set during the provisioning process and are transferred between the verifier and the verifier as part of the initialization step. The flow of security is beyond the scope of this document. By default, a step of $= 30$ seconds is suggested as a balance between security and usability.

The algorithms described above set the foundation for the development of one-time passwords for secure authentication.

Consider the model of the proposed information protection algorithm based on the random generation of a one-time key. Getting a one-time numeric key occurs according to the formula:

$$\text{Pass} = S(L, P)$$  \hspace{1cm} (8)

where: $\text{Pass}$ is the result of the calculated numeric key; $S$ – function for calculating a one-time numeric key based on user identification by the entered parameters $L$ and $P$; $L$ – user login; $P$ – user password.

We open the function $S(\cdot)$ as follows:

$$S(L, P) = T[i, j] \begin{bmatrix} a, b, c, x, y, p1, p2 \end{bmatrix}$$  \hspace{1cm} (9)

where: $T[\cdot]$ is an array of functions for calculating a one-time numeric key;

$$T = \begin{bmatrix} t_{i1} & t_{i2} & \cdots & t_{ij} \\ t_{i1} & t_{i2} & \cdots & t_{ij} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \cdots & t_{ij} \end{bmatrix}$$  \hspace{1cm} (10)

$i, j$ – position of functions in the array $T[\cdot]$; $a, b, c, x, y, p1, p2$ - function parameters for calculation from the array $T[\cdot]$. 
The position of the function in the array $T[\cdot]$ and its parameters are the result of obtaining a certain number from the generated hash. Find:

$$i = K(H, 1), j = K(H, 2)$$
$$a = K(H, 3), b = K(H, 4), c = K(H, 5),$$
$$x = K(H, 6), y = K(H, 7),$$
$$p_1 = K(H, 8), p_2 = K(H, 9)$$

where: $K(H, i)$ is the receiving function; $i$ - values from hash $H$

$$H = F(L, P, t, s)$$

where: $F(\cdot)$ is a hash calculation function based on input parameters:

$L, P, t, s$ - are input parameters,
$t$ – current system time/date;
$s$ – generated secret word.

Based on the described model, a two-factor authentication algorithm was developed.

IV. TWO-FACTOR AUTHENTICATION ALGORITHM

The proposed information security system using a combination of two factors: permanent and temporary passwords [5]. Permanent password (the first factor) the user chooses and uses when registering an account (account). A one-time or temporary password (the second factor) is generated on the server according to the proposed algorithm [5] and is valid for a specific period of time for one authentication session. The advantage of a one-time password is that the password is not reused. Thus, an attacker who intercepted data from a successful authentication session cannot use the copied password to gain access to the protected system. Temporary password generation is possible online. This password is generated on the server and displayed to the user in additional software on the smartphone with a validity period of ~20 seconds. The temporary password is generated based on the result of the selected function, which has a number of variable parameters. The function is combined into a table with a dimension of 256x256 multiple of a power of two. A detailed description of the operation of the algorithm is presented in [6].

The choice of this function and its initial parameters is based on the result of the hash function of the SHA256 standards [7-9]. This is a cryptographic hash function that was developed by the US National Security Agency.

Most modern systems for identifying, authorizing and protecting information are based on the use of the hash calculation function. To implement the system in question, we used the SHA 256 function, which calculates the hash based on the input string.

The SHA256 function for calculating the hash was used in view of its features that changing one bit of the input string affects all bits of the calculated hash, which eliminates the calculated dependence of the received hash on the input string.

The purpose of the hash function is to convert (or hash) an arbitrary set of data elements into a fixed-length value. This value will characterize the set of source data without the possibility of extraction.

The input string for the hash function is a combination of user credentials, the current time in GMT and an additional secret string. The result of the hash function is divided into separate numbers, which will be the indexes for the function selection and its initial data. The secret string is a required field, which will be selected from the array randomly. In connection with the analysis, it was decided to develop a generator that allows you to randomly generate words for choosing a secret word. Dictionaries of words were not used, since words are easier to crack. The generator is based on the use of the Latin alphabet of uppercase and lowercase characters, equal to 52. The length of the generated word is 5 characters. To analyze the generator, the exhaustive search method was used. According to this method, the line length is taken into account (the application shows the line length of 5 characters) and for example, the search speed is 100,000 words per second, we get $525 = 38,020,403,2$, resistance to 26 bits, the search time is 63 minutes. Due to the fact that according to the developed two-factor authentication algorithm, a one-time password is generated every 20 seconds, the probability of breaking the generated secret word is practically impossible. This confirms the efficiency of the proposed generator.

The received secret string data is used to select a function. For implementation, a function generator has been developed, the use of which will greatly facilitate their formation. To generate a function, the number of variables is taken as the basis. There are 7 of them in this generator. Initially, a list of variables is formed, as a result of which we obtain a random number of variables Count from 1 to the number of variables minus 1. Then we iterate through the array with certain variables N times based on a random number from 0 to the length of the array minus 1. A variable is read from the array, which is added to the new array and removed from the old one. After the cycle is completed, a list of variables for the function is generated. Based on this list, the constituent parts of the functions are formed. As a result, a list of components with variables is formed. Next, rows are formed based on a random number from 1 to 3. In the cycle, the components randomly merge, separated by signs of mathematical expressions. After receiving the strings, they merge separated by mathematical expressions. As a result, we obtain the generated string function, which we use to calculate the one-time password for two-factor authentication. A more detailed description of generators is presented in [10].

A. Information security model in an automated system based on the second factor

The described model and algorithm of the proposed method of protecting information in an automated control system using a combination of two factors: a permanent and temporary password. The user chooses a permanent password (the first factor) and uses it (account) when registering an account. The developed model is based on two types of two-factor authentication: authenticator applications and login verification using mobile applications are implemented in accordance with Figure 1.

Before automation, you must register in the application. After that, you need to run the application to enter user data (login and password), which must correspond to the registered data. If you successfully enter data, you must enter the application on your mobile device and enter the initial data to generate a temporary password. A temporary or one-time password (second factor) is generated on the server using the described algorithm and is valid for a specific period of time for one authentication session. The advantage of a one-time password is that the password cannot be reused.

Thus, an attacker who intercepted data from a successful authentication session cannot use the copied password to gain access to the protected system.
V. IMPLEMENTATION OF THE PROPOSED METHOD

To implement the model described above, the Node.js system is used. This software platform is responsible for writing the server side in the JavaScript programming language. Along with it comes the npm package manager, which is used to install various libraries and frameworks. Additionally, you must install MongoDB - an open source document management system for databases that does not require a description of the table layout. When complex queries arise, they are usually resolved on the application side, which makes it easier to work with data and links to them. The use of this database management system (DBMS) is due to the fact that rather simple scalability is built into this system using sharding technology, which splits (partitions) the database into separate parts so that each of them can be transferred to a separate server.

The advantage of using MongoDB DBMS is:
- increase in development speed;
- there is no need to synchronize the circuit in the database and application;
- clear path to scalability;
- simplicity of prescribed solutions.

If in relational databases the contents are tables, then in MongoDB the database consists of collections.

A collection is a group of MongoDB documents that is the equivalent of a simple table in a relational database. The collection is placed inside one database. A document in a collection can have various fields. Most often, all documents in the collection are created for one or related to each other purposes [11].

To work with the database, MongoDB was chosen because, unlike relational databases, it does not use a table device with a clearly defined number of columns and data types. A document in MongoDB can be represented as an object that stores some information. In a sense, it is similar to rows in relational DBMSs, where rows store information about a single item.

The choice of a document-oriented system was not accidental, since the MongoDB database, which stores login information and temporary information for authentication, uses an internal data encryption method based on a cryptographic algorithm. In the code of the developed application, the administrator himself sets the login and password:

```
MongoClient("mongodb+srv://admin:Mdb12812122424@cluster0-o83lo.mongodb.net /test?retryWrites=true", {
useNewUrlParser: true 
}).
```

This setting allows you to organize database protection. Database protection is an essential part of storing user data. Such results of hacker attacks as the leak of personal data and unauthorized access to information are the result of its insufficient protection. To determine the protection methods, you need to define the objects of protection. Such objects can be the values of certain fields, tables or records, the records themselves in tables, individual tables and entire databases. To protect these objects, you can use methods such as views, triggers, and built-in data encryption functions.
A view is to dynamically fetch data from multiple tables. It can also be called a virtual table, the formation of which depends on the user request directed to the available view. Using this approach, the user gets access only to the totality of columns and records that are determined by the used presented. Thus, you can limit the available data and control the set of information with edit access. Based on this, the presentation method can be used to protect the confidentiality and integrity of the data.

A trigger is a listener that is called during the execution of certain events. Such events are adding, deleting or editing data in the database. Using triggers, you can define user access for interacting with data in the database. They also allow you to record all events related to data modification in the protected table. Thus, this method must be used to effectively control data integrity and to monitor all events that occur during information changes.

Encryption features are not available in all database management systems. In MySQL, they provide data encryption using AES and DES algorithms. Hashing is also available using the MD5 and SHA1 algorithms. Using the available encryption features greatly simplifies the task of protecting data in databases.

Also important is the protection of the database from SQL injection. SQL injection is a common way to hack sites and applications. The purpose of such an attack is to disrupt the operation or obtain database data, without the right to access it. At the same time, injection, depending on the database management system used and implementation conditions, can allow an attacker to execute arbitrary database queries. An SQL injection type attack may be possible due to incorrect processing of the input used in SQL queries. Database application developers should be aware of such vulnerabilities and be able to prevent the possibility of SQL injection.

The MongoDB used in the current project is protected from SQL injection, since it is not a relational database management system of a document-oriented type and uses JSON ()-like documents and a database schema.

A. Configuration steps for the created client-server application

To work with Mongo, the “data” folder is created on the “C” drive, and the “db” folder is created in it. To install the packages necessary to run certain parts of the project (“Server”, “devschacht”, “dip”), you need to open the console in the folder with the package .json file and execute the “npminstall” command in it. After installing all the packages, the project is launched in parts. At the beginning, “Server” starts. To do this, go to the folder, open the console and execute the “npmstart” command. After starting, the message “server is running.” Next, the mobile application is launched through the “devschacht” command. After starting, a QR code will appear in the console. A tab with system information and a QR code will also open in the browser. You must remember the field with the IP address above the QR code.

For the mobile application to work, it is necessary to read the QR code, which can be done using the installed scanner on the mobile device by default or to install additional EXPO software. This is used to install the application. To start the smartphone and personal computer, the local network is set up so that they were on the same network. Next, the web site starts, the console opens, and the "npmrundev" command is executed. The web application uses port 4000 “localhost: 4000” to operate. To view the contents of the database, you need to install MongoDBCompass and make connections by clicking the “connect” button. The detailed description was presented in [12].

VI. RESULTS AND DISCUSSIONS

The study of existing algorithms and protocols for multifactor authentication was the beginning to create a modified user identification algorithm based on the second factor.

The developed system, based on the receipt of a one-time password, will allow the use of protection against unauthorized access in an automated control system, which is relevant. The use of security features for identification based on the second factor is used in almost all automated systems. Software implementation is described in [13].

The software implementation of the proposed method shows that the considered algorithm works correctly and corresponds to the beliefs described above.

REFERENCES