Analysis of Major Factors Preventing Cybercrime Reduction in Kazakhstan

Abdul Razaque[1], Hari M. Rai[1], Yersain Chinibayev[1] and Tolganay Chinibayeva[1]

[1]*International Information Technology University, Manas St. 34/1, Almaty, 050040, Kazakhstan*

Abstract

Abstract— Information systems are tightly integrated into the lives of modern people. They are also connected to cybercrime problems, which create a possible risk of losing personal confidential data and money. People and businesses must learn to follow simple rules to ensure information security. In this paper, we propose the necessary security measures to reduce cybercrime based on a set of applications and readily available solutions created using the Java language, which named Java Security Helper (JSH). The main objectives of this study are to review and analyze existing cybercrimes, identify the root causes of those cybercrimes, and create plausible security measures for both types of end-users (business, common internet users) with further implementation. The specific focus is on the development of innovative measures to combat cybercrime within the Republic of Kazakhstan.

Keywords

Cybercrime, Fighting cybercrime, High-tech crime, harm reduction, cyber literacy.

1. Introduction

It is evident today how dependent the world is on information technology. Ordinary users become targets of cyberattacks, and for many years they have donated their personal information and money to online criminals [1]. There are more and more chances in this area every day for enterprises as well as the general public in terms of quantity, quality, and accessibility. This demonstrates unequivocally that, given the increasing sophistication of information systems, vigorous marketing and sufficient security measures are needed to ensure that technology satisfies strict efficiency requirements [2-3]. Many information systems have already been firmly embedded in our daily lives. By neglecting the fundamental concerns that are impeding the decline of cybercrime in Kazakhstan, we risk not only being duped, losing our personal data, and our own savings, but we also make our loved ones an easy target for cyber fraudsters [4].

The introduction of cloud computing has also been a significant advancement in information systems. This not only enhanced processing power and speed of operations all across the world, but it also created new potential for cybercriminals. The country's poverty is a major contributor to the rise of many crimes, including cybercrime. A country's poverty can be defined in several ways: Low income, education, and health care are all common indications. Kazakhstan is an oil-based resource-based state with significant financial issues.

Some sectors of the population's financial weakness create a possible threat to the state on the Internet [5]. Cybercrime has increased dramatically in recent years all around the world. Business and enterprise networks are typically the primary targets of scammers. Cybercrime costs the global economy 445 billion dollars each year on average, with individual losses topping 160 billion dollars [6]. Understanding all of the threats to the economy, business, and individuals, users and business leaders must learn to adhere to simple standards to secure information security. Today, the government is attempting to create measures and laws to regulate

DTESI 2023: Proceedings of the 8th International Conference on Digital Technologies in Education, Science and Industry, December 06–07, 2023, Almaty, Kazakhstan

a.razaque@iitu.edu.kz (A. Razaque); r.hari@iitu.edu.kz (H.M. Rai); y.chinibayev@iitu.edu.kz (Y. Chinibayev); t.chinibayeva@iitu.edu.kz (T. Chinibayeva)

0000-0003-0409-3526 (A. Razaque); 0000-0001-7651-7338 (H.M. Rai); 0009-0009-8985-5892 (Y. Chinibayev); 0000-0002-2657-3697 (T. Chinibayeva)

cybercrime, but they fail to consider the ineffectiveness of such restrictions in modern circumstances.

The existing system employed by government organizations to record cyber-attacks is chaotic and wasteful, resulting in a huge amount of 143 million recorded attacks on government websites [7], implying that every single HTTP request or packet is counted as a distinct attack. Imposing exceptional sanctions is a common method used to tackle the country's expanding cybercrime. A government, for example, may contact an Internet service provider to restrict a harmful or viral Internet source. The state adopts special laws or programs to ensure that such rules are implemented more efficiently [8]. The legislature has been tasked with addressing the issue of compromised cyber security by establishing laws and launching state-level initiatives. Laws and regulations governing domestic activities such as organizational, legal, technical, and educational activities were drafted and implemented.

The action plan for its implementation spans five years and focuses on the government recognizing prospective cyber-attacks and, if feasible, preventing them [9]. However, it should be highlighted that even after three years, this set of restrictions is not having the desired effect. According to cert.gov.kz data, the total number of cyber threats grew by 2% between 2018 and 2019 [10]. This is a top-down decision: the state enacts legislation that begins to affect corporations and regular citizens. However, because of the idiosyncrasies of our society and psyche, actions from on high do not always have the desired effect. Ukraine, as a CIS country, followed a similar route to Kazakhstan.The initial stage was to establish a specific organization to regulate the country's level of cyber security, followed by the passage of legislation and the expansion of the powers of law enforcement [11]. In the early years, Russia followed a similar path, but one differentiating aspect was the establishment of harsh punishments for lawbreakers [12]. My solution is based on reverse reasoning. Actions will be taken "bottom-up": from the level of an average citizen and user to the level of the state. As a result, the general public's level of education in the subject of cyber security will rise, which will benefit existing small and medium-sized firms.

The proposed solution is to provide a solution that will aid in increasing the self-awareness of ordinary people in the sphere of cybersecurity. The "bottom-up" methodology will be used to implement this collection of ready-made IT solutions. The main contributions of this paper are given as:

- Determination of the current level of cyber literacy skills of citizens or employees of Kazakhstani companies through the use of surveys and saving statistics in the application.
- Potential increase in the level of knowledge of users of the Kazakhstani segment of the Internet in the field of cyber security.
- Provision of a basic set of tools for training employees of Kazakhstani companies in the field of cyber security (not related to the IT department).

The rest of the paper is organized as follows:

Section II identifies the problem and its significance. Section III discusses the related work of the existing method. Section IV discusses the proposed system model of the current work.

2. Problem identification and significance

One of Kazakhstan's major concerns today is the poor degree of cyber literacy among all sectors of the population. Our country is only now beginning the active stage of IT technology development and is learning about cybersecurity issues. However, this does not change the fact that a lack of cyber literacy places major constraints on our country's ability to reduce cyber dangers. As previously said, there are issues with cyber literacy in all segments of our people, beginning with ordinary citizens using the Internet, moving on to enterprises with their IT goods and the worldwide market, and concluding with the state with its programs targeted at the Republic of Kazakhstan's IT development.

This illiteracy has significant consequences such as the loss of critical private user data, the inaccessibility of local services over the Internet, financial losses, and massive losses of large data sets of Kazakhstan inhabitants. The government has already taken steps to improve the situation in the country, such as the Cyber Shield of Kazakhstan program; some companies hold security meetings when signing an employment contract; and our Kazakhstani web applications encourage users to take basic cyber security precautions, such as not telling anyone your password. However, the decisions have little or no influence and are merely for show. There is a need for a publically available solution that can quickly and easily determine whether an action observed on the Internet is hazardous to the user. This application or combination of applications will present the user with all of the information required for decision-making, as well as remind and supply the user with material to improve the user's cyber literacy.

Effective data visualization improves decision-making by allowing data analysts and decision-makers to quickly absorb information, identify anomalies, and communicate their findings to a wider audience. It also helps with storytelling, helping convey the narrative hidden in the data. Analyzing data without visualization can be like traveling in the dark, but with it, analysts can illuminate the path to valuable information and informed choices.

2.1. Related work

The salient features of existing methods are discussed in this section. Kazakhstan first encountered the issue of cybersecurity late in its development when compared to other countries. From this, we can conclude that serious decisions in this area either have not yet been taken or have been borrowed from other countries, without regard to the specifics of our culture. One of the solutions chosen in Taiwan is the timely and active updating of legislative acts, as well as keeping the ISP records and log files for a certain period of time. A solution developed in Chunga et al. [13] will allow to quickly respond to new types of cyber fraud, as well as create a single fund to combat cyber threats on a par with more developed countries. However, this does not resolve the issue in countries with less strict laws and also puts the personal data of users at risk.

Research conducted by Islam et al. [14] analyzes the application of a set of measures against cyber fraudsters in the UK. In this case, cybersecurity is viewed as a dynamic phenomenon and a great attention is given to human-related risks and concepts related to social psychology and cultural development. This solution follows a hybrid top-down and bottom-up approach that combines theory and data-driven analysis.

Dupont et al. [15] researched creation of private companies own combined teams of technicians and lawyers to fight cybercrime. The creation of their own centers for investigation of cyber-attacks in large private companies will significantly speed up and improve the quality of these very investigations in the country.

One potential solution developed by Dupont et al. is the use of specialized software and applications for collecting statistics, such as the Cyber Readiness Index (CRI) [16]. This allows the relevant organizations to conduct a more accurate and detailed analysis of cyber fraud, but it does not help to reduce it in any way, only simply accelerates the process of disclosing cyber-attacks that have already occurred against a particular enterprise or country.

Foros [17] suggested creation of national cyber police, which would include units of the Ministry of Internal Affairs of Ukraine, ensuring the training and functioning of highly qualified specialists both in the field of law enforcement and in the field of the latest technologies. This approach would result in a timely public notice about the emergence of new cybercrimes and effective interaction with similar law enforcement agencies in other countries. However, the first advantage can be nullified by a potential problem of information warfare and cyber attackers spreading misinformation.

One of the simplest and most obvious solutions is to promote a culture of cyber literacy among the country's population, proposed by Zems et al. [18]. Advice and solutions are given for both the average Internet user and businesses and government. This method allows the education of the population on the Internet. However, the main disadvantage of the presented solution is the lack of a detailed description of the solution and its implementation in modern society.

The existing approaches focus on providing simple learning method for different institutions. On the other hand, the proposed method can greatly be beneficial for adaptive learning systems that can improve education in two ways: personalized learning and dynamic concepts. Artificial intelligence approaches could also be utilized for adaption management in the field of cybersecurity.

2.2. System model

Building a model, special attention should be focused on fault tolerance. Since the application is expected to be used throughout the country, an important development step is to create a distributed, easily extensible prototype. The entire database will be located on one local machine (server). However, this database will be represented not by one MySQL process, but by a series of small clusters (replicas). Each replica is wrapped in a Docker container. The web server itself will contain the SQL balancer. Depending on the type of request (Insert, Update, Delete, or Select), the balancer will route to the appropriate database cluster. The Master cluster will receive Insert, Update, Delete requests. Select requests will be directed to Slave clusters, which are replicas with different latency periods. Depending on the priority of the operation being performed (if the data is of high priority, then the select request goes to the master), the appropriate cluster is selected. One of the clusters does not participate in general load balancing but is a backup copy of the master. This architecture is shown in Figure 1.

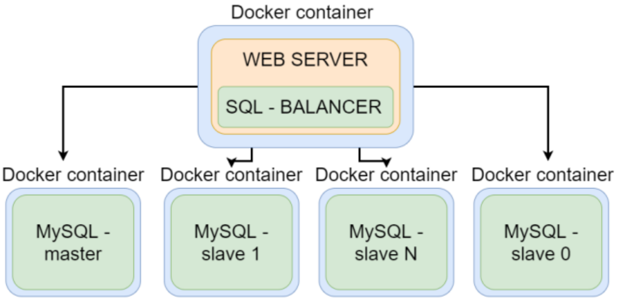

Figure 1: Database architecture (extendable)

Thus, we provide flexible SQL load balancing, the application itself is easier to administer; cheap and fast unlimited scaling is provided; the application is easier to run during the daily backup; one of our slaves, with a weight of 0, just serves for these actions and does not require updating the logic for large optimizations of the application itself.

The overall load of the application on the database is shown in equation (1).

$$\varphi = \sum_{i=0}^{n} C_s + C_m - C_{s0} \tag{1}$$

where φ is the total database load, C_s is the slave containers, C_m is the master container, C_{s0} is the zero weight slave container created exclusively for backups.

The general model of the entire architecture of the application is a mobile application that is installed by the end-user. All requests are sent to a single Backend server. The web server has a SQL balancer which interacts with our distributed database. Our database is an array of clusters, one master and the rest are slaves. When using images and other small media files, the system will access an external storage resource to unload our local server. For security measures uploading files to the server in any other way (for example, via FTP) will be prohibited. The API for working with images will include not only the binary data of the file, but also its metadata, such as size, title, and others. These parameters will be entered into the database for further retrieval and work with them. The general structure of the program is shown in Figure 2.

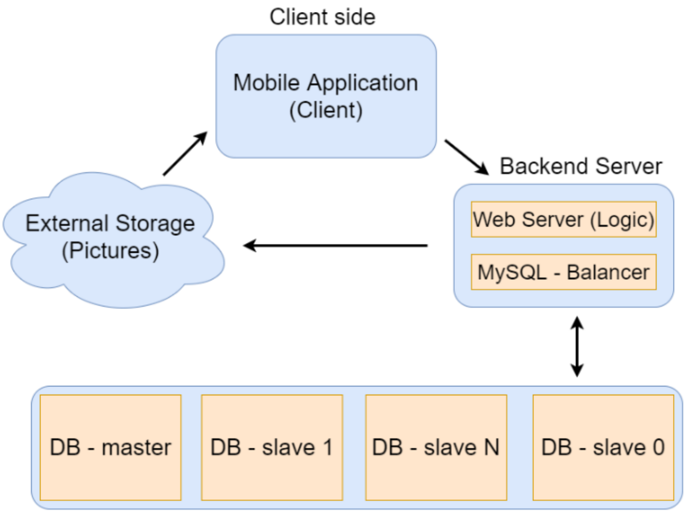

Figure 2: Application system architecture

Mobile application module. The mobile application module is responsible for displaying the results of processing information received from the backend module to the end-user. At the stage of prototyping, the development of this module will not be carried out.

External Storage module. External storage module - stores all types of media files, thereby relieving the backend application server. As a potential option, it is possible to use the Cloudinary service, since the main type of stored files belongs to the type of images (jpg, png, gif and others). Communication takes place through a backend server that stores information about the file: its metadata, for example, name, type file size. The backend server also stores a Uniform Resource Identifier (URI) for a potential file. All data is stored as cells in the database. When requested from the user, the server looks for this file in the database, receives it, and sends the necessary information to the mobile application, which in turn reproduces the URI from the external storage.

Backend server module. Backend server (module) - consists of two main parts: business layer and serving layer. The Business layer contains all the executable logic of the application. Requests from the mobile application come to the business layer and are processed according to the application's algorithms, then the necessary transformations and calculations are performed. If the result needs to be saved for further use, the business layer of the application transfers it to the serving layer.

The backend server maintenance layer is a set of solutions for communicating with various parts of the project. It includes connectors to remote external storage, database connections, and support for created singleton objects within the application. The serving layer performs mostly low-level tasks such as transferring data to the database, updating data in the database, or deleting objects from the application memory. It also houses the MySQL balancer, which evenly divides database queries between containers. More important requests with higher priority are sent to the master container (when it comes to reading). In other cases - for Slave-N containers.

Database Module. The database module stores information for which changes are required in the future. To reduce the load on the common backend server (the database is located on the same server), it was decided to reduce the load by dividing the database into equivalent containers. One container is a master container, all requests related to adding, updating, deleting are sent here. Other containers are replicas of the master container with different time delays. Read requests are made through these containers to unload the main master. One container has zero weight (0-slave), no read operations are performed on it, and its delay is about 24 hours. It is required to create a backup database in case of unforeseen circumstances.

3. Proposed cybercrime reduction process

The proposed solution is a set of ready-made services built on the Java programming language to improve the ordinary user's cybersecurity expertise. This program will be able to automatically notify the user in order to keep the user's cyber literacy at an adequate level. This algorithm will examine (via integration services) the user's most recent actions, and on the basis of this, ready-made solutions at the level of global security practices will be supplied. This solution employs a novel strategy for Kazakhstan. Kazakhstan has used state regulations to implement a top-down approach to cybersecurity in recent years. The current law on "Cyber Shield of Kazakhstan" is a noteworthy illustration. The developed solution uses a bottom-up methodology: starting from ordinary citizens of our country, passing through different layers of Kazakhstani business, ending with state-owned companies. This method of combating cyber illiteracy is a needed addition to the existing policies in the Republic of Kazakhstan. The work process of the application to improve the cyber literacy of the population consists of a set of utilities that interact with each other in turn. The application backend is split into several sections:

- Data preparation;
- User Decision handler.

Data preparation

This step helps to prepare and analyze the data that is given in algorithm 1.

```
Algorithm 1:  Data preparation to analyze
Initialization: { $A_U$ : User Action, $R_U$ : User Request}
Input: {$A_U$ }
Output: {$R_u$ }
Set $A_u$
Set $R_u \leftarrow A_u$
```

In algorithm 1 Data preparation for future analyze is discussed. In step-1, initialization process of given variables is explained. In steps 2-3, input and output are shown respectively. Step-4 shows the forming of raw user data. Step-5 shows the process of converting user action data to future server requests.

User Decision handler

This step helps students to create the list of security advice that is provided in algorithm 2.

```
Algorithm 2:  Creation a list of security advice and suggestions
Initialization: {$S_L$ : Potential suggestion list, $V_L$ : Vulnerability list, $K_D$ : Knowledge database, $R_U$ : User Request, $F_A$ : Array of filters, $R_A$ : Response from application}
Input: {$R_U$ }
Output: {$R_A$ }
For $R_u$ = 0 to $F_A$= $F_A$ $size$
    If $R_u == F_A$ && $R_u$ ! $\epsilon$ $K_D$ then
        $V_L = V_L + 1$
    End if
    $S_L$ = $S_L$ + 1
End for
If $V_L \neq empty$ then
    Set $R_A \leftarrow V_L$
End if
Set $R_A \leftarrow S_L$
```

In algorithm 2 Creation a list of security advice and suggestions are discussed. In step-1, initialization process of given variables is explained. In steps 2-3, input and output are shown respectively. Step-4 starts checking current request information in all possible chained filters. Steps 5-6 show that if data is found in a specific filter, but does not exist in the knowledge database of good practices then it adds it to the possible Vulnerability list. Step-8 shows the adding of a new suggestion to the potential suggestion list. In Steps 11-12 algorithm checks whether the vulnerability list is empty or not, and if something is present then it adds this data to the final response to the user. Step-13 adds the last suggestion list to respond, and sends it to the user.

$$R_A(R_u) = \frac{R_U}{K_D} + \frac{R_U}{F_A} \geq 1 \rightarrow V_L \geq 1, R_U \in A_U \tag{2}$$

where

R_A: User Request;

R_u: User Request;

K_D: Knowledge database;

F_A: Array of filters;

V_L: Vulnerability list;

A_U: User Action.

Equation (2) shows the dependence. When the sum of the ratio of the user's request to the general information of the data and the ratio of the user's request to the security filters is greater than or equal to one, this indicates that there is a security problem in the data transmitted to the users and the number of potential vulnerabilities is greater or equal to one. This will be the result of the application's response to the user's request. It is also worth noting that a user's request can only be formulated with the help of some user action.

Hypothesis 1: Improvement in the level of cyber literacy of the population leads to a decrease in cybercrime in the country.

Proof: Today, for the successful conduct of business and the successful development of the country, an important thing is the development of policies to increase the security of the enterprise. There are different types of policies that target different areas of work. In this case, we should focus on cybersecurity policies. The main threat in this area is cyber-attacks. Since it is necessary to somehow calculate the potential danger of this or that attack on a particular enterprise, the term cyber risk is introduced.

Definition 1: Cyber risk is the likelihood of disclosure or potential loss as a result of a cyber-attack or data breach. When thinking about cyber risk levels, it is important to first consider vectors that can be used to compromise sensitive assets and the various devices and applications that are at risk of being compromised. Since it is very difficult to say in what units cyber risk is measured, the concept of baselining is used, a deviation up or down. Due to this, cyber risk can be expressed by the following equation:

$$R = P * C \tag{3}$$

where

R: Cyber risk;

P: Probability of attack;

C: Consequence of Attack.

It is important to understand that in this case the risk is calculated without taking into account any external factors, both opposing and accompanying. The full equation of cyber risk, or rather the residual risk, is calculated as follows:

$$R_r = \frac{t}{v} * P_o * I - C_e \tag{4}$$

where

R_r: Residual risk;

t: Number of threats;

v:Number of vulnerabilities;

P_o: Possibility of occurrence;

I: Power of impact;

C_e: Control of effectiveness.

Definition 2: Control effectiveness (CE) represents the total effectiveness of all the controls that act upon a particular risk. This includes those controls that affect the likelihood of the risk. Also called preventive controls. Consists of the following components: state defensive policies, Enterprise Cybersecurity Policies, and each employee's cyber literacy knowledge. CE can be represented by using the equation below:

$$C_e = C_p + G_p + \sum E_p \tag{5}$$

where

C_p: Company cyber risk policies

G_p: Government cyber risk policies

E_p: User cyber literacy knowledge

It is important to understand that the cyber literacy of not only the employees themselves but also the cyber literacy of people who develop protective policies must also be at a high level. Thus, the company will use not only the knowledge of its employees to counter cyber-attacks, but also various preliminary measures. For example, ordering pentest audits, bringing your services to the platform's bounty bug to increase the overall security level, and can be expressed by the equation:

$$C_p = \sum(\delta + \alpha + E'_p), \qquad E'_p \in E_p \tag{6}$$

where

δ: One of bug bounty systems;

α: One of the performed audits;

E'_p: One employee, with cyber literacy.

This equation shows the dependence of all actions aimed at improving the level of security and reducing the potential number of cyber threats. The δ value is the number of bug bounty systems related to the web services of this country or company. The α value accounts the number of audits performed for critical services domestically. The E'_p value refers to the cyber literacy of a single employee. The knowledge of any person has a very strong effect on the overall level of cybersecurity, not only in a single enterprise, but in the entire state as a whole, thus the following equation can be obtained:

$$E_p \uparrow \rightarrow C_e \uparrow\uparrow \tag{7}$$

Based on the data obtained, the higher the cyber literacy of each person in the state, the lower the level of risk of one or another threat. It is important to note that it is not only the cyber literacy of individual users or employees but also cyber knowledge in the field of the safety of people who develop both company security policies and laws for the state. From here we can get the following equation:

$$C_e \uparrow \rightarrow R \downarrow \tag{8}$$

Corollary 1: Thus, by expanding the various areas of cybersecurity, we reduce the overall likelihood of a major cyber-attack with enormous consequences. Responsible for expanding the level of cybersecurity is the cyber literate population at all levels: from the average user to people who make state laws.

Hypothesis 2: The more people find new similar cybersecurity threats with a switched-on application, the less likely it is that this threat will result in harm to the user in the future.

Proof: It is a known fact that over time, new threats appear in the field of IT services and products. It is not always possible to respond to emerging threats in a timely manner. One of the obvious solutions is to minimize the damage that can be received from the newly-minted cyber threat as much as possible. Ordinary users and employees of small companies often become the main targets of cybercriminals. To prevent this problem, one of the potential solutions could be this project. It is logical that a system cannot be ready for all possible (or not yet existing) types of attacks on a user, however, receiving similar requests from users, it is easy to find a relationship and define a new type of threat, and then notify all users about this new cyber threat and build a course of action against it. All existing cyber threats can be expressed in the following equation:

$$W = \sum (T_k + T_z + T_f) \tag{9}$$

where

W: All cyber threats;

T_k: Cyber threats or cyber-attacks that are already well studied and exist in many bases;

T_z: Cyber threats, that have appeared relatively recently, the impact of which is not fully known;

T_f: Potential cyber threats that do not yet exist, but they will form in the future.

The knowledge database of this proposal contains information on a number of pre-existing cyberattacks, which it actively monitors and updates. Also, relatively new types of attacks are also found in this database. The content of the database is indicated in the equation below:

$$K_D = T_k + T_{z\prime} \tag{10}$$

where

$T_{z\prime}$: Cyber threats, that have appeared relatively recently, the impact of which is not fully known and exist in knowledge database.

It is worth noting that the value of T_z from Equation (9) does not equal the value of $T_{z\prime}$ from Equation (10). However, they are in a certain dependence on each other, which is shown in the following equation:

$$T_z \neq T_{z'}, T_z \,!\in K_D, T_{z'} \in K_D, T_{z'} \in T_z \tag{11}$$

The $T_{z'}$ value is an element from the T_z value array and is contained in the application database. Thus, it is easy to determine that not all cyber-attacks of this type are contained in the application system. The next step is to define the client base of the application for subsequent calculations. The equation for finding the percentage of users who use this application is shown below:

$$P = \frac{U_o}{U_a} \tag{12}$$

where

P: The percentage of users who use this application;

U_o: Users, who belong to application system;

U_a: All internet users of Kazakhstan.

The number of all users of the Kazakhstani segment of the Internet is taken from statistical data in the public domain from the datareportal [19], Internet resource for 2020. The number of users within the system is calculated using the following equation:

$$U_o = U_r \pm \Delta U \tag{13}$$

where

U_r: Registered user in the system;

ΔU: Error for inactive users who use the application very rarely.

$$\Delta U = U_r - U_{act} \tag{14}$$

where

U_{act}: Active users.

The value of active users is found by tracking the average daily number of users using Grafana Zabbix. After finding the coefficient of the user of the application system, you can determine how dangerous the new threat is and how often it has been encountered recently. If we imagine that P_t is the coefficient of a new cyber threat, we can find it using the following equation:

$$P_t = \frac{P * N}{100} \tag{15}$$

where

N: The number of requests related to this particular cyber threat.

After finding the severity ratio of a given cyber threat, you need to compare it with any value to determine how high the priority of this cyber-attack is and whether it is subject to further research. The result can be seen in the equation below:

$$P_t \geq 0.35 \ \rightarrow P_t \in R \tag{16}$$

where

R: Array of investigated cyber attacks.

If P_t exceeds the value of 0.35, then this problem is assigned the "Growing" status, which means that it is potentially dangerous in the future and the problem is then sent for investigation. The coefficient of 0.35 is not chosen by chance and is based on research by Kaspersky Lab in 2017 [20].

The next step after investigating a new cyber threat is to add it to the database of the project. This action is described by Equation (17):

$$K_D \ \leftarrow \ P_t, P_t \ \in \ T_z \,\&\, P_t \ \in \ T_{z'} \tag{17}$$

After adding a new threat, possible scenarios for preventing this threat are added to the database. After completing these operations, the update is added to the released version of the application. Thus, users with the installed application and with the next message about this threat already receive a ready-made solution or instructions for further actions. Thus, the number of users who can become potential victims of the new cyber threat is reduced. The “word of mouth” effect is also useful. It lies in the fact that users who are not familiar with this cyber threat and do not use the application will be prepared and ready to take action against this threat. Thus, we can express the number of people who can successfully confront this problem with the following equation:

$$U_s = \ U_o + U_d - \ \Delta U' \tag{18}$$

where

U_s: Users who will be able to fight the cyber threat;

U_d: Users who heard about the problem from users in the U_o category;

$\Delta U'$: Error associated with external factors, the prevalence of the problem among the public.

Using the above calculations, we can determine the propagation coefficient of a new cyber threat in the following equation:

$$P_c = \frac{N_t}{U_s} \tag{19}$$

where

P_c: Cyber threat propagation coefficient;

N_t: The number of meetings of this cyber threat.

Thus, it can be seen that the more users become aware of and know how to fight a new cyber threat, the lower the prevalence rate of this cyber threat is, potentially approaching zero. It is important to understand that the time taken to obtain this result is not taken into account here. Time periods can greatly affect the end result of a given equation.

Corollary 2: The more people encountered and found this vulnerability, the more people are ready for a similar type of attack. It follows that the preparedness of people affects the chance that a new cyber threat will somehow affect them. And this directly depends on how much lower the damage will be, both to an individual citizen, as an individual enterprise, and to the whole country as a whole

Theorem 1: A company's negligent attitude towards potential cyber threats from the outside leads to serious financial losses and further ruin of the company.

Proof: Today, small and medium-sized businesses in Kazakhstan are still in a developing state. Very often, this particular financial sector becomes a potential attack site for cyber criminals.

In order to determine the financial losses of a company, we will introduce the ROI index.

Definition 1: ROI determines how much a company earns in comparison with the amount spent. The indicator is expressed as a percentage of the returned investment over a specific amount of time. ROI equals the present value of net benefits accumulated over a certain time period divided by the initial costs of investment. ROI is calculated using the following equation:

$$ROI = \frac{B - C_i}{C_i} \tag{20}$$

where

B: Benefits;

C_i: Cost of investment.

Definition 2: Benefits are the advantage or profit gained from something. It is very difficult to define the benefit by any one equation. Therefore, in our case, the benefit will be represented as a difference between Annual Loss Expectancy (ALE) without security investment and ALE with security investment (16).

$$B = A_t - A_w \tag{21}$$

where

A_t: ALE without investment;

A_w: ALE with investment.

ALE represents the total amount; which an organization could lose in one year if nothing is done to mitigate the risk. ALE can be calculated by the following equation:

$$A = S * A_r \tag{22}$$

where

S: Single Loss Exposure;

A_r: Annual Rate of Occurrence.

Single Loss Exposure (SLE) is the total amount of money, which lost from one risk that happened. Consist of the product of Asset Value and Exposure Factor and shown in equation below:

$$S = A_V * E_F \tag{23}$$

where

A_V: Asset Value;

E_F: Exposure Factor.

This is how the ROI can be found. However, it should be noted that one of the factors is not considered. The formula contains potential risks for the company, but there is no account for negligent attitude of employees as related to the safety of the enterprise. To indicate this value, an A factor will be used to indicate an employee's negligence about cyber security within the company. Then, equation (20) is supplemented with the following fragment:

$$ROI_A = ROI - n * H \tag{24}$$

where

H: Negligence rate;

n: Number of such employees.

On Equation (24), not only the risks that are associated with the company's product and market are dangerous for the business itself, but every employee who is not educated on cyber literacy also exposes the entire enterprise to additional risks, as well as directly affects the potential income of the company.

4. Experimental results

To evaluate the performance of the proposed approach, test conditions based on a high load service written in Java were built. The Spring helper framework was chosen to increase code performance and speed up development, and it was launched using Spring Boot. The database was built using the MySQL database management system. Another optimization step was to place many containers on the same server. Containers were created using Docker. Table 1 describes all calculations conducted on a local PC with parameters as near to the combat server as possible.

Table 1: Hardware characteristics

Operating System	Ubuntu 18.04.5 LTS (Bionic Beaver)
CPU	Intel® Core™ i5-9400F CPU @ 2.90Ghz (6 CPUs)
Processor Architecture	x64
Random Access Memory (RAM)	16384MB
Paging File	11609 MB
GPU	NVIDIA GeForce GTX 950, 2GB
Hard Drive	HDD 1Tb, SSD 256 Gb

Further, Table 2 contains all the characteristics of the versions and parameters of the developed application.

Table 2: Software characteristics

Java version	8
Installed JDK	OpenJDK Amazon Coretto
Spring Boot version	2.4.0
MySQL server version	5.7.18 Community
Number of Docker container	3

Based on testing, the following parameters are evaluated to determine the effectiveness of the proposed idea: Fault Tolerance, Percentages of Threats, Cyber Security knowledge.

A. Fault Tolerance

The Scenario-1: Test the fault tolerance in high loads:

One of the crucial processes to test is the fault tolerance of an application. There are 2 potential weak points: polling and availability of external storage of images and communication of the backend server with the database. Since the first vulnerability directly depends on the availability of this service as such and the speed of the Internet connection, this issue was not considered.

To simulate the load on the server, third-party services that created a huge number of requests and called endpoints that performed operations with the database were used. To resemble realistic conditions, the requests were sent more actively and in larger quantities in the afternoon, between 3 pm and 9 pm.

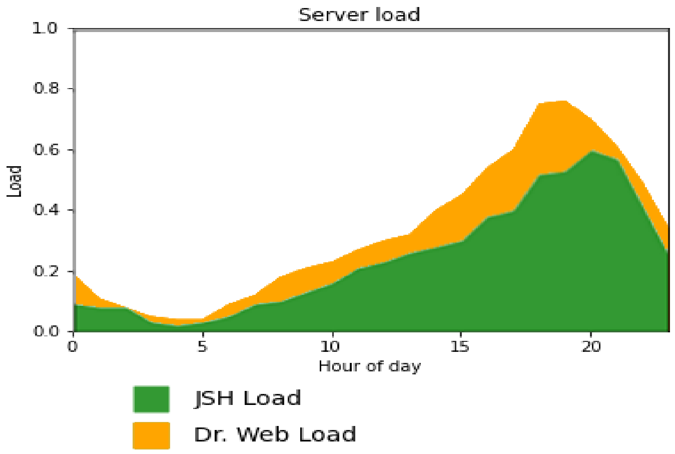

Figure 4: Comparison of Backend server load

To achieve the best performance result, the following number of containers were taken: 1 "master" container, 1 "zero" container and 2 "slave" containers to achieve the maximum possible performance improvement without investing too much money in supporting this architecture. The JSH system was compared with an existing solution on the market: Dr. Web for servers. Server load results using two different solutions over a full day are shown in Figure 4.

Based on the results obtained, it can be seen that the used JSH application consumes fewer resources allocated by the server, which will have a positive effect on the speed and performance of the entire server as a whole, which directly affects the speed at which the user receives information. When comparing the peak loads on the server, you can see that the JSH loads during the busiest period by 17% less than the other consumer solution.

JSH is extensible, but the effect with the same number of requests with an increase in the number of containers will be smaller. The decrease in performance with an increase in the number of containers can be expressed by the following equation:

$$f(x) = -\frac{1}{2x+1} + 1, x \in [0;\ +\infty] \tag{25}$$

where x is the number of database containers.

The result of the increase in productivity can be clearly seen in Figure 5. Thus, it can be seen that the real increase is felt when a small number of containers is added.

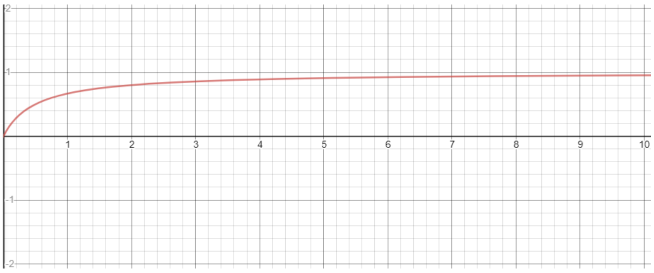

Figure 5: The server performance with increasing number of database containers

B. Percentages of Threats

The scenario-2: Comparison of User Assistance Tools in Combating Cyber Threats

This scenario compares our proposed approach with other existing the state-of-the-art approaches: Mobile antivirus Doctor Web [21] and a reference manual mobile application [22]. The comparison was carried out according to the following criteria: The percentage of blocked threats in the joint of the user plus the enabled application, and the metric by which the results were submitted is the level of threats by the user himself. The last check is the result of what the user has learned using this application. The results of the experiment can be found in Figure 6.

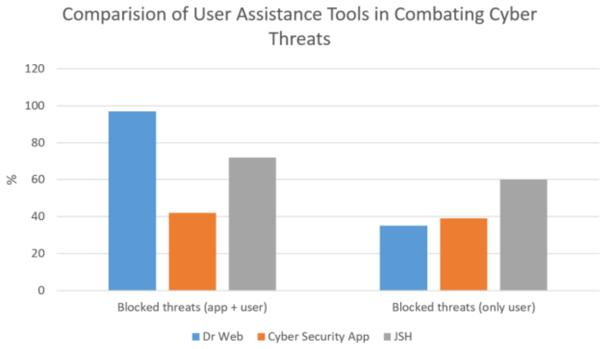

Figure 6: The results of comparison different applications

It is worth noting that the check on the latter basis took place after active use of the application for a week. Thus, the results showed the following. An antivirus is a good solution, but only when it actively works in the user's system, without it, the user shows weak knowledge and is a potential target for cybercriminals. The directory application has weak indicators, both with the application actively turned on and without it. The user has to study the information in the manual himself, which is why his level of cyber literacy is very low. The solution provided in the project does not show good results immediately. It is designed for long term use. Its main task is to educate the user in cyber literacy and prepare this user to face possible cyber dangers.

C. Fault Tolerance

A control group was recruited to conduct a survey and test among the employees of the "Kredit 24" company. 10 employees unrelated to the IT and information security department were allocated. Three employees belong to the tele-marketing department, three from the business module department, three call center operators and one from the legal department. A basic knowledge survey related to cyber literacy was conducted. The results of the primary survey showed a very low result. 6 participants answered correctly less than 50%, only 2 people gave a satisfactory result, which is above 75% correctly answered questions, after the survey. 50-75% correctly answered questions were considered a weak result. The overall test results can be seen in Figure 7.

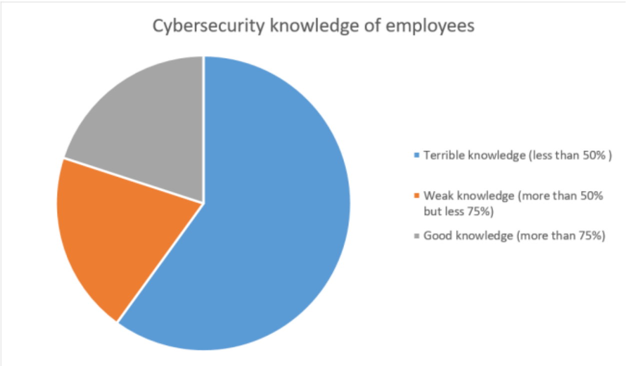

Figure 7: The result knowledge in the field of cybersecurity

A second survey was then conducted after users had used the proposed application for a period of one week. The results are displayed in Figure 8.

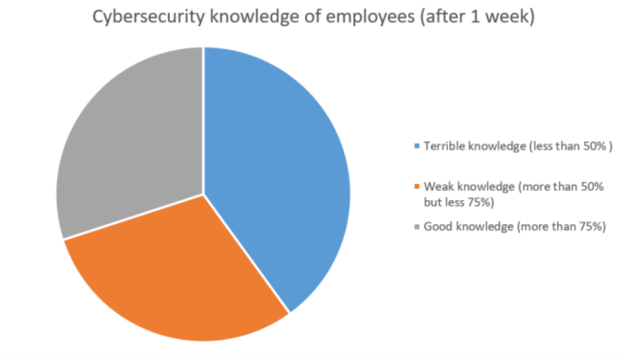

Figure 8: Employee cybersecurity knowledge after 1 week

5. Discussion of results

The proposed solution was tested with server and client load, the number of threats detected with and without utilizing the application. Furthermore, it was compared to other solutions, the proposed application demonstrated stable performance under high loads during the day. This conclusion is greatly reliant on the number of containers deployed and how well the load is distributed among them.

The undeniable advantage is that all calculations and operations are performed directly on the server, without affecting the user device, preventing the final program from being loaded and thereby improving the user experience. The performance with the most favorable number of containers deployed was 17% better than the other approach. In the second experiment, the number of potential cyber threats blocked by the user was compared to the number of blockings during active use of the application with the user's knowledge. The designed application performed mediocrely in both active protection (application plus user) and autonomous work by the user.

The reason for this is that it is not aimed at an independent search and prevention of cyber risks, but rather at assisting the user in determining if this is a potential hazard or not. The program loses 3% to the antivirus Dr. Web because it lacks active security features. The application was then tested on real individuals to see how well it teaches people who are not directly tied to IT but utilize IT products. Users were shown an unfinished prototype of the application. One of the benefits is the potential decrease in risks for small, medium, and large firms in the country. This is due to the expanded cybersecurity experience of all corporate personnel, not just those in the information security or development departments. The second significant benefit is the widespread adoption of cyber literacy among the general public. The biggest issue with such a solution is the lack of actual physical application among the people or within any firm, which might produce a significant variation from theoretical study results. The speed with which potential users can be acquired is also a significant consideration. This method is many times slower and takes considerably longer to get the desired result than the top-down method, which runs in a second. The test results show a good trend: two participants moved from the group that answered fewer than 50% of the questions correctly to the group that answered less than 75% correctly. One person progressed from the "weak" to the "good" group. Thus, a qualitative rise in the cyber literacy of the company's personnel of 30% can be observed in a relatively short time of 7 days.

6. Conclusion

This article introduces an application to prevent cybercrime in Kazakhstan. The proposed application provides a collection of security advice and proposals to assist in the prevention of cybercrime in Kazakhstan. The application integrates several emerging technologies to ensure cybercrime prevention through the adaptive learning process. The main advantage of the proposed application is to use of the Spring Framework backend server. The framework enables developers to modify the program without incurring additional financial or time expenditures. The framework also provides high robustness to manage large server loads and provides versatile built-in capabilities to fulfill developers' demands. Furthermore, an additional development stage is included for building the SQL balancer, which enables query optimization and downtime minimization. Cloudinary is employed as an external data storage system that allows to alter images on the fly without requiring the primary backend server to be loaded. Tests have been conducted that demonstrate effective fault tolerance, prevention of cyber threats, and improving security knowledge. Furthermore, the testing results indicate that the proposed solution improves the subject knowledge of users from 20% to 30%. The cyber literacy has also increased from 40% to 60%. Even if this is not ideal, it will be taken into account in future work to deal with contemporary problems. Artificial intelligence will be incorporated with the proposed solution to increase the prevention ratio of cybercrime in Kazakhstan. The application will fully be integrated into a Kazakhstani company.

7. References

[1] Razaque, A., Al Ajlan, A., Melaoune, N., Alotaibi, M., Alotaibi, B., Dias, I., ... & Zhao, C. (2021). Avoidance of cybersecurity threats with the deployment of a web-based blockchain-enabled cybersecurity awareness system. Applied Sciences, 11(17), 7880.

[2] Impacts of information technology, March 21, 2018. URL: https://master-iesc-angers.com/impacts-of-information-technology-it/.

[3] Homan Forouzan, Hamid Jahankhani, John McCarthy (2018). An Examination into the Level of Training, Education and Awareness Among Frontline Police Officers in Tackling Cybercrime Within the Metropolitan Police Service. Advanced Sciences and Technologies for Security Applications, pp. 307-323.

[4] Aliya Tabassum, Mohammad Saleh Mustafa, Sumaya Ali Al Maadeed (2018). The Need for a Global Response Against Cybercrime. 6th International Symposium on Digital Forensic and Security, pp. 1-6.

[5] Razaque, A., Kejun, D., Xueqi, Z., Wanyue, L., Hani, Q. B., & Khan, M. J. (2018). Survey: Wildlife trade and related criminal activities over the internet. In 2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1-6.

[6] The impact of cybercrime on small business. URL: https://www.sbir.gov/tutorials/cyber-security/tutorial-1#.

[7] Razaque, A., Alotaibi, B., Alotaibi, M., Hussain, S., Alotaibi, A., & Jotsov, V. (2022). Clickbait detection using deep recurrent neural network. Applied Sciences, 12(1), 504.

[8] Emilio C. Viano (2017). Cybercrime: Definition, Typology, and Criminalization. Cybercrime, Organized Crime, and Societal Responses, pp. 3-22.

[9] Isabaeva Symbat, Botagoz M. Yesseniyazova (2019). Cyber security issues in digital Kazakhstan.

[10] Incident statistics by years, 2020. URL: https://cert.gov.kz/press_club/infographics.

[11] Borko, Andrii, Vadym Nehodchenko, Olena Volobuieva, Ivan Kharaberiush, and Y. S. Lohvynenko (2019). Fighting against cybercrime: problems and prospects in Ukraine and the world.

[12] Klimovskikh, J., V. Tsapko, and O. Gridina (2018). The problem of cybercrime in Russia. In International scientific research, pp. 103-104.

[13] Macas, M., Wu, C., & Fuertes, W. (2023). Adversarial examples: A survey of attacks and defenses in deep learning-enabled cybersecurity systems. Expert Systems with Applications, 122223.

[14] Tasmina Islam, Ingolf Becker, Rebecca Posner, Paul Ekblom, Michael McGuire, Herv´e Borrion Shujun Li. (2019). A Socio-Technical and Co-Evolutionary Framework for Reducing Human-Related Risks in Cyber Security and Cybercrime Ecosystems. Dependability in Sensor, Cloud, and Big Data Systems and Applications book.

[15] Benoit Dupont (2016). Bots, cops, and corporations: on the limits of enforcement and the promise of polycentric regulation as a way to control large-scale cybercrime. Springer Science+Business Media Dordrecht.

[16] Benoît Dupont (2019). Enhancing the effectiveness of cybercrime prevention through policy monitoring. Journal of Crime and Justice, pp. 500-515.

[17] Sarkar, G., & Shukla, S. K. (2023). Behavioral Analysis of Cybercrime: Paving the Way for Effective Policing Strategies. Journal of Economic Criminology, 100034.

[18] Zems, Mathias (2015). Techniques for Cybercrime Prevention and Detection.

[19] Digital 2020, Kazakhstan, February 18, 2020. URL: https://datareportal.com/reports/digital-2020-kazakhstan.

[20] The Human Factor in IT Security, 2017. URL: https://www.kaspersky.com/blog/the-human-factor-in-it-security/.

[21] Dr. Web Security for Mobile. URL: https://products.drweb.com/mobile/android/?lng=en.

[22] Porcedda, M. G. (2023). Sentencing data-driven cybercrime. How data crime with cascading effects is tackled by UK courts. Computer Law & Security Review, 48, 105793.