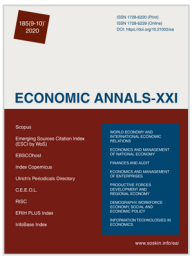

ECONOMIC ANNALS-XXI
ISSN 1728-6239 (Online)
ISSN 1728-6220 (Print)
https://doi.org/10.21003/ea
http://www.soskin.info/ea/

Volume 185 Issue (9-10)'2020

Citation information: Nelub, V., Gantimurov, A., & Borodulin, A. (2020). Economic analysis of data protection in systems with complex architecture using neural network methods. *Economic Annals-XXI, 185*(9-10), 178-188. doi: https://doi.org/10.21003/ea.V185-17

UDC: 332.05

Vladimir Nelub
PhD (Engineering),
Bauman Moscow State Technical University;
Director, Interdisciplinary Engineering Centre «Composites of Russia» of the Bauman Moscow State Technical University
1 Lefortovskaya Embankment, Moscow, 105005, Russian Federation
mail@emtc.ru
ORCID ID:
https://orcid.org/0000-0001-6712-7135

Andrey Gantimurov
PhD Student (Economics),
Bauman Moscow State Technical University;
Chief Technology Officer,
BAUM-Inform
1 Lefortovskaya Embankment, Moscow, 105005, Russian Federation
a.gantimurov@npobaum.ru
ORCID ID:
https://orcid.org/0000-0002-5000-702X

Alexey Borodulin
PhD (Engineering),
Bauman Moscow State Technical University;
Deputy Director, Interdisciplinary Engineering Centre «Composites of Russia» of the Bauman Moscow State Technical University
1 Lefortovskaya Embankment, Moscow, 105005, Russian Federation
asb@emtc.ru
ORCID ID:
https://orcid.org/0000-0003-4448-6762

Economic analysis of data protection in systems with complex architecture using neural network methods

Abstract

Introduction. In the United States, Europe, and Asia, there have been spikes in cyber attacks on protected and confidential information (including bank data, personal data, and confidential business information) over the period 2000-2020.

Data protection in systems with complex architectures is a complex and non-trivial solution, which is suitable for flexible self-tuning and self-learning tools, such as neural networks as state modern studies.

The described above state of things stipulates the importance and topicality of the direction of our research. Our research raises the question about the economic study of the data but attempts to create a mathematical apparatus to cause a loss of data in digital form, not made before.

The purpose of the study is to propose and test the calculation of the economic value of the data loss and economic benefit from data protection with multi-level neural networks.

Results. Nowadays, almost every business that has access to the Internet has a threat of losing protected data, including banking data. To prevent this, the company must comply with the data protection management regulations. As an example, we studied the process of neural network reaction to an attack and its detection, when two types of attacks are threatened: botnet and UDP flood. Due to the result of a quick response to detect suspicious activity, neural network methods are excellent for use in economic analysis, because the neural network logs every event for every millisecond.

Conclusions. As a result of using neural network tools for economic analysis of data protection in modern systems with complex architecture, we were able to obtain stable results of responding to an emerging attack in order to obtain protected data. Such protection is a preventive measure for the possibility of reducing business losses in the event of data loss. As we can see, the management of protection of systems with complex architecture is necessary for each company with the specified data level.

Keywords: Information Security; Data Protection; Neural Network; Economic Analysis; Cyber Attack
JEL Classification: C15; L78
Acknowledgements and Funding: The authors received no direct funding for this research.
Contribution: The authors contributed equally to this work.
DOI: https://doi.org/10.21003/ea.V185-17

Нелюб В. О.
кандидат технічних наук, Московський державний технічний університет імені Н. Е. Баумана;
директор, Міжгалузевий інжиніринговий центр «Композити Росії» МДТУ ім. Н. Е. Баумана,
Москва, Російська Федерація
Гантімуров А. П.
здобувач наукового ступеня кандидата економічних наук, кафедра промислової логістики,
Московський державний технічний університет імені Н. Е. Баумана;
головний технічний директор, ТОВ «БАУМ-Інформ», Москва, Російська Федерація
Бородулін О. С.
кандидат технічних наук, Московський державний технічний університет імені Н. Е. Баумана;
заступник директора, Міжгалузевий інжиніринговий центр «Композити Росії» МДТУ ім. Н. Е. Баумана,
Москва, Російська Федерація

Економічний аналіз захисту даних у системах зі складною архітектурою методами нейронних мереж

Анотація

Вступ. Економічна вартість захисту даних, особливо якщо розглядати складноструктурні банківські й захищені дані, надмірно висока. Одна зі значущих атак за 2020 рік була здійснена на Поштовий банк Японії, з клієнтських рахунків якого були вкрадені 570000 доларів США. Репутаційні ризики банку прорахувати при цьому не видається можливим. Із ростом кіберзлочинності в світі значення безпеки, актуальності й складності використання захищених даних стає все більшим із кожним роком. За 2020 рік не тільки банки зіткнулися з кіберзлочинністю, але також і соціальні мережі (Twitter, Reddit), великі компанії (AT & T, Templatemonster), мережеві ритейлери (Wallmart, Tesco) і фірми в інших сферах підприємницької активності по всьому світу.

У США, Європі та в країнах Азії за період 2000–2019 років відзначаються сплески кібернетичних атак на захищену й конфіденційну інформацію (зокрема банківські дані, особисті дані, конфіденційна бізнес-інформація).

Захист даних у системах зі складною архітектурою є складним і нетривіальним рішенням, для якого підходить гнучкий самоналагоджувальний і самонавчальний інструментарій, яким і є нейронні мережі, що підтверджується в сучасних дослідженнях.

Сьогодні стоїть питання про економічне дослідження даних, однак спроб для створення математичного апарату, що дозволяє виразити втрати даних у фінансовій формі, не робилося.

Мета дослідження полягає в пропозиції та апробації розрахунку економічної вартості втрати й захисту даних за допомогою багаторівневих нейронних мереж.

Результати. Нині практично кожен бізнес, який має вихід в Інтернет, має загрозу втрати захищених даних, у тому числі й банківських. Для запобігання цьому в компанії повинен дотримуватися регламент менеджменту захисту даних. Як приклад ми досліджували процес реакції нейронної мережі на кібератаку та її виявлення при загрозі двох типів атак: ботнет і UDP флуд. Завдяки результату швидкої реакції виявлення підозрілої активності методи нейронних мереж відмінно підходять для використання в економічному аналізі, позаяк нейромережа протоколює кожну подію за кожну мілісекунду.

Висновки. У результаті застосування інструментів нейронних мереж для економічного аналізу захисту даних у сучасних системах зі складною архітектурою нам вдалося отримати стійкі результати реагування на атаку з метою отримання захищених даних. Такий захист є превентивним заходом для можливості скорочення збитків бізнесу при втраті даних. Як ми бачимо, менеджмент захисту систем зі складною архітектурою є необхідним для кожної компанії зі вказаним рівнем даних.

Ключові слова: безпека інформації; нейронна мережа; захист даних; економічний аналіз; кібераттака; кіберзлочинність.

Нелюб В. А.
кандидат технических наук,
Московский государственный технический университет имени Н. Э. Баумана;
директор, Межотраслевой инжиниринговый центр «Композиты России» МГТУ им. Н. Э. Баумана,
Москва, Российская Федерация
Гантимуров А. П.
соискатель ученой степени кандидата экономических наук, кафедра промышленной логистики,
Московский государственный технический университет имени Н. Э. Баумана;
главный технический директор, ООО «БАУМ-Информ», Москва, Российская Федерация
Бородулин А. С.
кандидат технических наук,
Московский государственный технический университет имени Н. Э. Баумана;
заместитель директора, Межотраслевой инжиниринговый центр «Композиты России»
МГТУ им. Н. Э. Баумана, Москва, Российская Федерация

Экономический анализ защиты данных в системах со сложной архитектурой методами нейронных сетей

Аннотация
Вступление. Экономическая стоимость защиты данных, особенно если рассматривать сложноструктурные банковские и защищенные данные, чрезмерно высока. Одна из значимых атак за 2020 год была на Почтовый банк Японии, с клиентских счетов которого были украдены 570000 долларов США. Репутационные риски банка просчитать при этом не представляется возможным. С ростом киберпреступности в мире значение безопасности, актуальности и сложности использования защищенных данных становится все больше с каждым годом. За 2020 год не только банки столкнулись с киберпреступностью, но также и социальные сети (Twitter, Reddit), крупные компании (AT&T, Templatemonster), сетевые ритейлеры (Wallmart, Tesco) и фирмы в других сферах предпринимательской активности по всему миру.
В США, Европе и в странах Азии за период 2000–2019 годов отмечаются всплески кибернетических атак на защищаемую и конфиденциальную информацию (в числе которой банковские данные, личные данные, конфиденциальная бизнес-информация).
Защита данных в системах со сложной архитектурой является сложным и нетривиальным решением, для которого подходит гибкий самонастраивающийся и самообучающийся инструментарий, каким и являются нейронные сети, что подтверждается в современных исследованиях.
Сегодня поднимается вопрос об экономическом исследовании данных, однако попыток для создания математического аппарата, позволяющего выразить потери данных в финансовой форме, не предпринималось.
Цель исследования заключается в предложении и апробации расчета экономической стоимости потери и защиты данных с помощью многоуровневых нейронных сетей.
Результаты. На текущий период времени, практически каждый бизнес, имеющий выход в Интернет, имеет угрозу потери защищенных данных, в том числе и банковских. Для предотвращения этого в компании должен соблюдаться регламент менеджмента защиты данных. В качестве примера мы исследовали процесс реакции нейронной сети на кибератаку и её выявление при угрозе двух типов атак: ботнет и UDP флуд. Благодаря результату быстрой реакции выявления подозрительной активности методы нейронных сетей отлично подходят для использования в экономическом анализе, так как нейросеть протоколирует каждое событие за каждую миллисекунду.
Выводы. В результате применения инструментов нейронных сетей для экономического анализа защиты данных в современных системах со сложной архитектурой нам удалось получить устойчивые результаты реагирования на возникающую атаку с целью получения защищенных данных. Такая защита является превентивной мерой для возможности сокращения убытков бизнеса при потере данных. Как мы видим, менеджмент защиты систем со сложной архитектурой является необходимым для каждой компании с указанным уровнем данных.
Ключевые слова: безопасность информации; нейронная сеть; защита данных; экономический анализ; кибераттака; киберпреступность.

1. Introduction

The economic cost of data protection, especially when considering complex banking and secure data, is excessively high. One of the most significant attacks in the year 2020 was on the Postal Bank of Japan, from whose client accounts USD 570 000 was stolen; the bank's reputational risks cannot be calculated. With the rise of cybercrime in the world, the security, relevance and complexity of using secure data is becoming more relevant every year. In 2020, not only banks faced cybercrime, but also social networks (Twitter, Reddit), large companies (AT&T, Templatemonster), chain retailers (Wallmart, Tesco) and other areas of business activity around the world.

Data protection in systems with complex architectures is a complex and non-trivial solution, which is suitable for flexible self-tuning and self-learning tools, such as neural networks (Ucci et al., 2017; Souri et al., 2018; Sultana et al., 2019; Arief et al., 2015; Qiu et al., 2016; McLeod et al., 2018; Algarni et al., 2016; Kafali et al., 2017; Sen et al., 2015; Ablon et al., 2016).

Our research raises the question about the economic study of the data but attempts to create a mathematical apparatus to cause a loss of data in digital form, not made before (Edwards et al., 2015; Abelson et al., 2015; Anderson et al., 2010; Romanosky et al., 2017; Ruffle et al., 2014; Shu et al., 2017; Gordon et al., 2015; Biener et al., 2015).

2. Brief Literature Review

The combination of AI and certain technical systems equipped with actuators and sensors is called «robot», which has long been widely used (Avigur-Eshel, 2018). The term «robot» has become even more popular recently, which is also facilitated by the fact that in recent decades Robotics has received huge development due to achievements in nano- and biotechnology, microminiaturization of technical devices, in computer technology and programming, in Big data

Processing (big data) and much more (Bawden et al., 2008). A particularly impressive boost to robotics has been given by the fact that AI technologies have become much more efficient and diverse in recent years, and, most importantly, by the fact that they have become economically easily accessible (Campbell, 2012; Carr, 2016; Chang et al., 2018).

Almost all large technological corporations in all countries are actively developing or already implementing «robots» designed for use in a wide variety of human activities related to both physical and intellectual actions or, in other words, related to the implementation of certain actions with both material and non-material content of objects (de Bruijn et al., 2017). This perception of the concept of «robot» is very important for the problems of the Internet of things, since it is robots that allow you to perform various services and perform work due to the presence of various actuators (Fitzgerald, 2016). It is quite obvious that in the context of the use of IP technologies, robotization becomes necessary for objectifying in the real world the results of the implementation of certain cognitive functions by artificial intelligence (Hadjimatheou, 2019).

So, in order for the results of AI functioning to be used (consumed) by humans, it is necessary to transform them with the help of robots into some objects that have a material or non-material meaning (Shrouf et al., 2015).

Nevertheless, to this day, among a wide range of specialists and scientists, including lawyers, the science fiction and Multimedia perception of robots still prevails. Therefore, back in September 2019, experts from the world of robotics and other fields of knowledge stated: We need to figure out how best to integrate robots into the social, legal and cultural framework of our society, how to take into account the interests of people from different cultural traditions, who will not look at our work with a wide range of assumptions, myths and stories behind them (Zhu et al., 2017).

3. Purpose

The purpose of the study is to propose and test the calculation of the economic value of the data loss and economic benefit from data protection with multi-level neural networks.

4. Methods

In our research we use methods of the theory of automatic control - management, optimization, and object-oriented programming for data. The main provisions of the work are obtained on the basis of reliable knowledge of applied mathematics with mathematically rigorous calculations. The obtained theoretical results are confirmed by computational methods' experiments.

5. Results

5.1. Composition of a Typical Attack for Obtaining Protected Information

Violations of the level of security in the use of electronic computing tools, telecommunications systems and computer networks are divided into the following types:

- unauthorized access to the operation of automated systems, computer networks, databases;
- creation for the purpose of using, distributing or marketing malicious software products or technical means, as well as their distribution or sale;
- unauthorized sale or distribution of restricted information stored in automated systems, computer networks, or on such information carriers;
- crimes committed by using a computer system as a means of achieving a criminal goal, etc.

Attacks are carried out by attackers to violate the privacy, integrity, or availability of holes stored, processed, and distributed in its. Let's take a closer look at some of these vulnerability databases:

- CVE-help and search engine for links and vulnerability designations. An important component of the system is CVE identifiers - unique identifiers provided to each known information security vulnerability. The system contains more than 98,000 records of individual vulnerabilities.

The main difference is that it is the most complete and systematic, so it is used as a basis for determining whether vulnerability records match other databases. The main elements of the vulnerability record structure include:

1) status-this field can contain either The Entry Value (verified record) or the value (vulnerability not yet tested);
2) phase - this field contains the value of the vulnerability development stage, as well as the date of assignment of the specified stage. There can be the following values:

- proposed - the vulnerability suggestion phase;
- interim - the intermediate phase of the vulnerability;
- modified - the vulnerability modification phase;
- assigned - the vulnerability installation phase;

3) description - the field contains a description of the vulnerability;
4) links - this field contains links to other sources with a specific address of the internet resource describing the vulnerability and source ID;
5) votes - the field contains the names of voting members who decided to enter the vulnerability in the database;
6) comments - enter the name of the author of the comment and its text content.

The vulnerability record elements also contain the vulnerability type, name, and ID. The vulnerability name is in the format «CVE-YYYYNNNN», where YYYY is the year of vulnerability discovery, and NNNN is its serial number.

The process of adding a vulnerability to the database consists of three steps:
1) processing - Analysis, Research, and the process of bringing vulnerabilities to the CVE format;
2) assignment - assigning a CVE identifier to a specific vulnerability record;
3) publication - create a new entry and publish it on the CVE internet resource as soon as the CVE identifier is officially assigned.

The disadvantages of the CVE database include the lack of a mechanism for describing whether vulnerabilities belong to specific products, as well as assigning metrics to vulnerabilities and calculating the degree of danger.

The US National Vulnerability Database (NVD) is a repository of vulnerability data based on the standards of the security content automation protocol. The NVD database combines a description of vulnerabilities, software names with these vulnerabilities, and vulnerability risk assessments. As of 2020, the NVD Vulnerability Database had about 70,000 vulnerability records.

The structure of the vulnerability record in the NVD database is an extended form for presenting a record in the CVE database, due to the presence of the following fields: configuration of vulnerable products based on dependencies; list of vulnerable products; indicators that characterize the vulnerability in the format of the «general vulnerability assessment system»; access type for implementing the vulnerability. It was found that 82.77% of vulnerabilities belong to applications, 12.28% to operating systems, and 3.59% to hardware.

A distinctive feature of the NVD database is the use of Common Platform Enumeration, which is one of the best product dictionaries among well-known analogues due to the large number of entries and the unified format of software and hardware names. However, this format for presenting product records has drawbacks, namely:

- ambiguity of values of different format fields;
- insufficient use of entries in this dictionary by the NVD database.

The OSVDB vulnerability database is designed for a community of security professionals. The goal of the project is to provide accurate, detailed, up-to-date information about vulnerabilities for security systems. This database contains more than 110,000 vulnerabilities.

The most common attacks are fragmented UDP Flood and Botnet attack.

Fragmented UDP Flood. This attack is carried out by fragmented UDP packets of small size, for the analysis and compilation of which the platform has to allocate resources. Deep traffic analysis systems also provide protection against this type of flood, discarding protocols that are irrelevant for the client site or limiting them by Lane. For example, for websites, the working protocols are HTTP and HTTPS. In this case, you can simply exclude or restrict irrelevant protocols by band.

Botnet attack. Attackers usually try to flood the victim's lane with a large number of packets or connections, overloading network equipment. Such large-scale attacks are carried out using many compromised systems that are part of the botnet (Figure 1).

In this example (Figure 1), an attacker controls several «zombie machines» to conduct attacks. «Zombies» communicate with the main machine via a secure hidden channel, and control is often carried out via IRC, P2P networks, and even via Twitter.

When performing this type of attack, the user does not need to hide the IP address of each machine, and due to the large number of computers involved in the attack, such actions lead to a significant load on the site. Moreover, attackers usually choose the most resource-intensive requests.

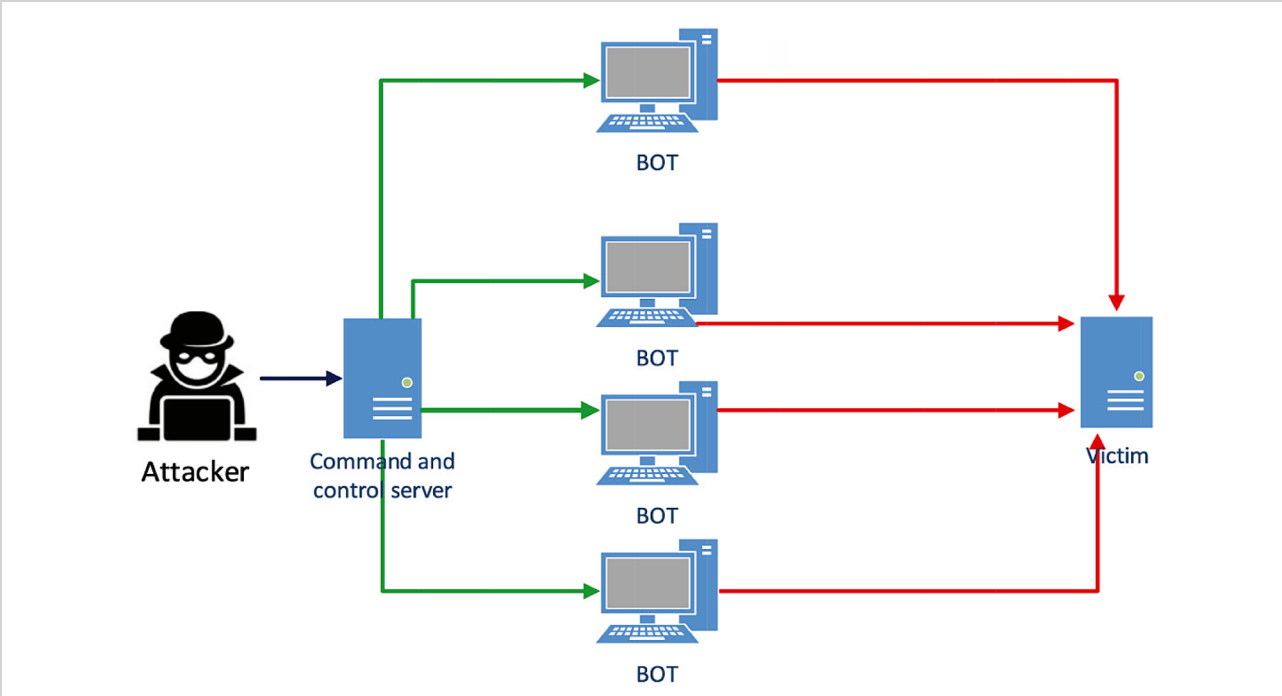

Figure 1:
Botnet attack
Source: Compiled by the authors

5.2. Information Security Model in the Complex Systems

The specification of the selected formal model that implements the basic concept of a mathematical algorithm consists of two points. The first point of the model specification is to maintain a new basic essence in the definition of an actor - a set of checks, which can determine the permissible time limits for calculations and the time for forwarding messages, as well as provide additional opportunities for evaluating the correctness of the actor's work. The second part of the specification includes introducing into the message forwarding control component of the computing system the procedure for running a set of checks for an actor, which is performed after its main functions are performed. It should be noted that the changes made do not affect the internal logic of the mechanisms of functioning of the mathematical information protection algorithm. Formulas of the research are collected in Table 1.

Let us take a closer look at the first part of the data specification.

Definition 1 (Actor, transition). Let U be the set of all possible values (universe), and Formula 1 be the set of all finite sequences in U.

For any nonempty set of states Σ, the actor $m \to n$ with sending is called the set Formula 2 initial state, Formula 3 transition relation.

An element with τ is called a transition. An element with A is called a check (assert). For any transition (5), the correct entry is (6) in which the state $\sigma(\sigma')$ is called the precursor (successor) of the state $\sigma'(\sigma)$, and $s(s')$ is the input (output) message. Then the set of actors $m \to n$ is defined as $Am \to n$, and the set of all actors is defined as Formula 7.

The introduction of a set of α checks into the actor definition is the first part of the MA specification, which provides the actor with the ability to check time and algorithmic conditions and constraints. To provide the necessary time analysis functionality, an incoming and outgoing message $s(s')$ must contain information about the time it was sent and received, i.e. $s = s \circ t$, where s - original message, $t \in S$ - tag (timestamp), $\circ$ - composition ratio.

Next, we will consider the second part of the specification, which is related to the computing system and defines the procedure for checking the set of asserts of a certain actor. A number of definitions are pre-entered.

Definition 2 (events, signals). An event is an element of a set, where T is the set of all labels, and V is the set of all values. A signal $s \in S$ is a set of events, where S is the set of all subsets of E. A tuple of N signals is denoted by N'.

An empty λ signal is a no-event signal. A tuple of N empty signals is denoted by ΛN.

Table 1:
Formulas of Research

$S = U^*$	(1)
$< \sigma_0, \Sigma, \tau, A >$, where $\sigma_0 \in \Sigma$	(2)
$A = \{\alpha 1 \ldots \alpha n\} \mid ai = pi\,(s, s'), p\,(s, s'): Sm \times Sn \longrightarrow \{0,1\}, s \in S, s' \in S, n \in N.$	(3)
$< \sigma_0, \Sigma, \tau, A >$, where $\sigma_0 \in \Sigma$	(4)
(σ, s, s ', σ') ∈ τ	(5)
$\sigma \dfrac{s \to s'}{t} \sigma'$	(6)
$\mathcal{A} = \cup m, n \in N\, \mathcal{A}m \longrightarrow n$	(7)
$s = [s_1, \ldots, s_N] \in S^N$	(8)
$T\,(s) \subseteq T$	(9)
$G = \{f: S^m \times T \longrightarrow S^n \times G\}$	(10)
$f\,(\Lambda^M, t) = f\,(\Lambda^N, f)$	(11)
$asrt: S^m \times T \times S^n \times T \longrightarrow S^k$	(12)
$\mathcal{A} = \cup m, n \in N\, \mathcal{A}m \longrightarrow n$	(13)

1. while $(s \neq \Lambda^N)$ {
2. let $\tau = min\,(T\,(s)))$
3. while $(s\,(\tau) \neq \Lambda^N)$ {
a. let $j = min\,\{k \in \{1, 2, \ldots, N\}: s_k\,(\tau) \neq \lambda\}$
b. let $i \in \{1, 2, \ldots, M\}: j \in Ii$
c. let $(s'', f_i) = f_i\,(\pi_{Ii}\,(s\,(\tau)), \tau)$
d. let $(s''') = as_i\,(\pi_{Ii}\,(s\,(\tau)), s'', \tau, \tau')$
e. let $s = s - Select_i\,(s\,(\tau)) \cup Expand_i\,(s'') \cup Expand_i\,(s''')$
4.}}
	(14)

$\mathcal{M}\,(OSI_{level}, char, t_{comp}, \delta T): N \times N \to \mathcal{U}, where\ OSI_{level}$	(15)
$char \in Chars = \{char_1, \ldots, char_p\}, p \in \mathrm{N}$	(16)
$t_{comp}, N = \{n_1, \ldots, n_k\}, k \in \mathrm{N}$	(17)
$\mathcal{D}\,(\mathcal{M}_1, \mathcal{M}_2) = \mathcal{M}_1 - \mathcal{M}_2$, where $dij = \mid \mu ij - \mu' \mid \in \mathcal{D}$	(18)
where $\mu ij \in \mathcal{M}_1, \mu' \in \mathcal{M}_2, i, j \in \mathbb{N}$	(19)
$NetCrits = \{netCrit_1, \ldots, netCrit_i\}, i \in \mathbb{N}.$	(20)
$f = \Sigma_{ij}\,(w_{ij} \times d_{ij})$, where $\Sigma_{ij}\, d_{ij}\, ij \in \{0,1\}$	(21)
$\varphi_1 = max_{NetCrits}\ netCrit\_i, Dom\ \varphi_1 = [0,1].$	(22)
$PlcCrits = \{plcCrit_1, \ldots, plcCrit_j\}, j \in \mathbb{N},$	(23)
where $plcCrit_j\,(t_{comp}, \delta T, t) = g_j\,(\mathcal{T}).$	(24)
$ALE_i = SLE_i \times ARO_i = (AV \times EF_i) \times ARO_i$	(25)
$ALE_i' = SLE_i \times ARO_i = (AV_i \times EF_i') \times ARO_i'$	(26)
$NPV(R, d) = \sum_{t=0}^{T} \frac{\sum_{i-1}^{l}(ALE_{it} - ALE_{it}')}{(1+d)^t}$	(27)

Source: Compiled by the authors

Definition 3 (individual labels). Set (10) is called a set of individual labels in the IR signal.

Definition 4 (transmission function). The transfer function f takes a set of input events and an individual label (time) and returns a set of output events and a new transfer function called continuation, moreover. Set of all transfer functions (11) for m inputs and IR outputs, with (12) for all $t \in T$.

Definition 5 (verification function). The validation function accepts sets of events and individual labels (times) and returns a set of output events

Operational semantics are defined for the interaction of actors. Let there be signals, actors with transmission functions $f1,..., fn$, empty sets of incoming messages $I_1,..., I_M$ and output $O_1,..., O_M$ indexes, $s \in S_N$ - the set of events present, then the operational semantics are defined as follows (formula (14) and below in Table 1):

The S signal represented in the loop definition on line 1 acts as a global message queue. The main loop continues until all events are processed. The variable τ in line 2 reflects the current time. Line 3 and the variable j defines the index of the signal with the lowest rank, containing input events at the time τ. In line b, the variable i defines the index of the process receiving the $S(\tau)$ signal as one of the input signals.

5.3. Concepts of the Reference Model Functioning

The functioning of the reference model is based on two criteria. The first criterion reflects the «correct» state (configuration) of the system. The second criterion simulates the «correct» operation of the system.

The ability to identify reference characteristics of the state and behaviour is due to the stability of the configured (within a certain technological process) SCADA system and the high frequency of traffic within its infrastructure, which is confirmed by the possibility of modelling data transmission channels (HMI-PLC) using its own finite deterministic Automata.

Thus, if there is information about the correct functioning of the periodic table in the past, it becomes possible to predict information about its correct functioning in the future. Different time intervals allow you to compare information within the work of a particular operator (hour), shifts (day), work schedule (month), seasonal changes (year), etc.

The criterion for the reference behaviour of the system is determined by the set of channels of interaction between PLC and HMI and is set using finite automata of these channels. The criterion for the reference state of the system is determined by two characteristics: the network state and the equipment settings (PLC).

For each criterion, a private security function is defined that evaluates the level of its compromise. The next step is to define a general security function that combines information received from private security functions.

To build the model, we will use network, statistical and diagnostic data obtained using Netflow protocols, as well as port mirroring methods. The complexity of constructing a reference model is not high due to the use of the mathematical apparatus of Fuzzy Logic and depends only on the number of elements of the network infrastructure. Setting up the model by an expert will consist in determining thresholds for functions that reflect the degree of deviation of the values of real parameters from the expected ones.

Thus, if the function value exceeds the safe threshold, a suspicious incident (virus software implementation) will be recorded in the model, and the management staff at the HMI station will be informed about this.

The criterion for the reference («correct») state (configuration) of the system is determined by a set of important system settings, parameters and properties characteristic of a properly functioning and non-compromised SU. The structural elements of the criterion are: network status and equipment settings (PLC).

To build a characteristic of the network State at the first stage, network activity matrices are formed from network Statistics data, reflecting the picture of information exchange at different levels of the OSI network model over a certain period of time. Matrices can contain information about the interaction of certain components (network adjacency matrices), statistical information about the total average amount of traffic transmitted, the time and period of interaction between specified nodes. Formally, The Matrix data is defined as follows, namely formula 15 - the selected level of the open systems model, formula 16 - the selected characteristic(the fact of interaction of network nodes, the average volume of traffic, etc.), t_{comp} - the time point from which the

construction of the matrix begins, δT - the time period starting from Formula 17 - the final set of network interaction elements at the $OSI\ level$ (for example, MAC addresses for the channel layer or a pair < IP address; port > for the transport layer), U - set of values of the network activity matrix (can be set {0,1} for adjacency matrices or R for the average traffic volume).

The column of this matrix corresponds to source nodes, and the row corresponds to destination nodes. The Matrix is square. To detect deviations in the work, two network activity matrices are compared, one of which is a reference one. In this case, a difference matrix of the form (18) is constructed.

If ij network matrices have different dimensions (this fact usually means the appearance of a new node or connection), the smaller matrix expands to a larger dimension, and the default values are used as values (for example, 0 for adjacency matrices). Columns and rows of matrices should be equally sorted.

Next, many functions will be defined-network status criteria (19)

Criteria (20), where t_{comp} - start of the reference measurement time, δT - reference measurement interval, t - start of a real measurement, D - set of difference matrices, M - set of network activity matrices, $Dom\ netCrit_i = [0.1]$.

Function f_i calculates the network compromise criterion based on the information contained in the network activity matrices (requires at least one matrix to calculate the criterion) with the construction of difference matrices D. Specific criteria from the set of NetCrits can be as follows (21).

Difference in adjacency matrices at the channel, network, transport layer, and Modbus level, taking into account the risk of deviations. In other words, formula 21 is an element of the difference Matrix corresponding to a given network level, and $w_{ij} \in [0,1]$ is the risk of this interaction ($\Sigma_{ij} w_{ij} = 1$).

The danger can be determined based on the classification of network elements by the degree of trust and the direction of interaction (for example, from an untrusted node to a trusted one and vice versa).

Query-Response matching at the Modbus level.

Matching the volume, time, and path tolerance (for example, to a PLC) of traffic. By combining the readings of various functions-network status criteria, the network surge function is set (22) to build a characteristic of the state of equipment settings at the first stage, tables of diagnostic data of controllers are formed.

Further, many criteria are formed from tabular data in the same way (23).

Comparable elements are vectors that contain information about controller parameters over a certain period of time. Specific functions-criteria for compromising hardware settings relate to the correct operation of the logic, configuration, RAM, and processor of controllers. An example is the function for estimating the number of running processes on the controller $g = \Sigma_i (w_i * d_i)$, where $w \in [0,1]$ - process flexibility (OS, production, $\Sigma_i d_i$ other), and $d \in \{0,1\}$ - difference in the vector for comparing running processes. The list of possible diagnostic data for analysis is determined by the capabilities of the equipment and protocols (for example, Diagnostic Subfunctions for Modbus).

5.4. Economic Calculation of the Effect of using a Neural Network to Protect Information

One of the few ways that can help determine the effect of the implementation of measures in the field of information protection in complex systems is a monetary assessment of the damage that can be caused to information resources, and which can be prevented as a result of the implementation of the proposed measures using neural networks. Thus, the expected damage prevented will constitute the resulting economic effect or additional cash flow. With this approach, most calculations can only be estimates and are approximate.

This is due to the fact that the activity of attackers, who are sources of threats to information security, is almost unpredictable: it is impossible to reliably predict the attack strategies, the qualifications of the attackers, their specific intentions and resources that will be used to commit certain actions, as well as intentions in relation to stolen information. Accordingly, in order to make all the necessary calculations, it is necessary to make a lot of assumptions and expert assessments in the context of the activities of this particular enterprise, as well as, if possible, to study statistical information concerning attacks on information resources similar to those protected.

Thus, the economic evaluation of the effectiveness of measures to protect the information system: an assessment of existing threats to information assets, which will affect the implementation of protective measures; assessment of the probability of each of the identified threats; economic impact assessment of the implementation of threats. It can be defined as the product of the total value of the protected information assets AV (Active Value) by the coefficient of their destruction due to the violation of information security EF_i (Exposure Factor). The number of information security violations for the year is Annual Rate of Occurrence (ARO_i) which is the estimated frequency with which information security violations of the i-th type occur during the year. The estimated value of the average annual loss is Annual Loss Expectation (ALE_i) which is the total amount of losses from information security violations (implementation of risks) of the i-th type during the year (24-25).

The immediate effect of the implementation of measures to improve the level of information security will be manifested in the fact that:

- the negative consequences of each implemented threat after the implementation of measures (EF_i') will be less than they were before their implementation: $EF_i > EF_i'$;
- the frequency of information security violations will decrease after the implementation of the $ARO_i > ARO_i'$ measures.

As a result, the reduced value of ALE_i' will be in (26).

Thus, the total annual effect from the implementation of the event will be defined as: Based on this, the total cash flow from the implementation of the event is determined by the following formula (27).

On the basis of these data, the total effect of the implementation of measures in the field of information security can be determined and it can be demonstrated how justified and appropriate investments in certain information security tools are in the conditions of a particular enterprise, taking into account all the features of its functioning. Although from a mathematical point of view, all the calculations in the described framework model for estimating ROI are simple, determining individual parameters can cause significant difficulties in practice. Conducting such calculations, as well as conducting information security audits, may require the involvement of third-party consultants, but the qualifications and professional specialization of such consultants may differ significantly from the qualifications of consultants specializing, for example, in conducting audits and implementing technical means of information protection.

6. Conclusion

Economic analysis of the security of information systems is quite rare in the scientific literature. In this paper, we study the system of neural networks for information security, and we came to the result of the system efficiency of 0.91.

The developed system has broad prospects for further development, as it is a component modular product and does not impose restrictions on the number of modules for analyzing various security aspects of information systems. The main prospects for the development of the system are: creating a module for assessing the trusted state after making changes, creating a module for analyzing and evaluating operator actions.

The main areas of further research are as follows. First, it is necessary to conduct a study to identify the features of placing and implementing a tracking service for the permissibility of operator actions on the human-machine interface (HMI). Secondly, it is necessary to conduct research on the creation of a generalized method for constructing a finite automaton of the PLC-HMI interaction channel. Third, it is necessary to conduct research on the features of the system's operation in conditions of processing large amounts of data (the so-called Big Data).

References

1. Abelson, R., & Goldstein, M. (2015, February 5). *Millions of Anthem Customers Targeted in Cyberattack.* The New York Times. http://www.nytimes.com/2015/02/05/business/hackers-breached-data-of-millions-insurer-says.html
2. Ablon, L., Heaton, P., Lavery, D., & Romanosky, S. (2016) *Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1187.html
3. Algarni, A. M., & Malaiya, Y. K. (2016). *A consolidated approach for estimation of data security breach costs*. In 2016 2nd International Conference on Information Management (ICIM) (pp. 26-39). IEEE. https://www.cs.colostate.edu/~malaiya/p/breachcostAlgarni2016.pdf

4. Anderson, E. T., Fong, N. M., Simester, D. I., & Tucker, C. E. (2010). How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet. *Journal of Marketing Research*, *47*(2), 229-39. https://doi.org/10.1509%2Fjmkr.47.2.229
5. Arief, B., Adzmi, M. A. B., & Gross, T. (2015). Understanding cybercrime from its stakeholders' perspectives: Part 1-attackers. *IEEE Security & Privacy, 13*(1), 71-76. https://doi.org/10.1109/MSP.2015.19
6. Avigur-Eshel, Amit. (2018). Synthesizing Depoliticization and Responsibilization: The Case of Financial Education in Israel. *Competition & Change, 22*(5), 509-528. https://doi.org/10.1177/1024529418798115
7. Bawden, D., & Robinson, L. (2008). The Dark Side of Information: Overload, Anxiety and Other Paradoxes and Pathologies. *Journal of Information Science, 35*(2), 180-191. https://doi.org/10.1177/0165551508095781
8. Biener, C., Eling, M., & Wirfs, J. H. (2015). Insurability of cyber risk: an empirical analysis. *The Geneva Papers on Risk and Insurance - Issues and Practice, 40,* 131-158. https://doi.org/10.1057/gpp.2014.19
9. Campbell, H. (2012). Planning to Change the World: Between Knowledge and Action Lies Synthesis. *Journal of Planning Education and Research, 32*(2), 135-146. https://doi.org/10.1177/0739456X11436347
10. Carr, M. (2016). Public-Private Partnerships in National Cyber-Security Strategies. *International Affairs*, *92*(1), 43-62. https://doi.org/10.1111/1468-2346.12504
11. Chang, L. Y. C., Zhong, L. Y., & Grabosky, P. N. (2018). Citizen Co-production of Cyber Security: Self-Help, Vigilantes, and Cybercrime. *Regulation & Governance*, *12*(1), 101-114. https://doi.org/10.1111/rego.12125
12. Cybersource. (2019). *2019 Global eCommerce Fraud Management Report.* https://www.cybersource.com/content/dam/documents/en/global-fraud-report-2019.pdf
13. de Bruijn, H., & Janssen, M. (2017). Building Cybersecurity Awareness: The Need for Evidenced-Based Framing Strategies. *Government Information Quarterly, 34*(1), 1-7. https://doi.org/10.1016/j.giq.2017.02.007
14. Edwards, B., Hofmeyr, S., & Forrest, S. (2015). *Hype and heavy tails: a closer look at data breaches.* Workshop on the Economics of Information Security. https://econinfosec.org/archive/weis2015/papers/WEIS_2015_edwards.pdf
15. Fitzgerald, C., & Cunningham, James, A. (2016). Inside the University Technology Transfer Office: Mission Statement Analysis. *Journal of Technology Transfer*, *41*(5), 1235-1246. https://doi.org/10.1007/s10961-015-9419-6
16. Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Zhou, L. (2015). Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity*, *1*(1), 3-17. https://doi.org/10.1093/cybsec/tyv011
17. Hadjimatheou, K. (2019, October 16). Citizen-Led Digital Policing and Democratic Norms: The Case of Self-Styled Paedophile Hunters. *Criminology & Criminal Justice*. https://doi.org/10.1177/1748895819880956
18. Kafali, O., Jones, J., Petruso, M., Williams, L., & Singh, M. P. (2017). *How good is a security policy against real breaches? A HIPAA case study.* In 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE) (pp. 530-540). IEEE. https://doi.org/10.1109/ICSE.2017.55
19. McLeod, A., & Dolezel, D. (2018). Cyber-analytics: Modeling factors associated with healthcare data breaches. *Decision Support Systems*, *108*, 57-68. https://doi.org/10.1016/j.dss.2018.02.007
20. Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing, 67,* 1-67. https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-016-0355-x
21. Romanosky, S., Ablon, l., Kuehn, A., & Jones, T. (2017). *Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?* Working Paper, RAND Justice, Infrastructure, and Environment. https://weis2017.econinfosec.org/wp-content/uploads/sites/3/2017/06/WEIS_2017_paper_28.pdf
22. Romanosky, S., Hoffman, D., & Acquisti, A. (2014). A empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, *11*(1), 74-104. https://doi.org/10.1111/jels.12035
23. Ruffle, S. J., Bowman, G., Caccioli, F., Coburn, A., Kelly, S., Leslie, B., & Ralph, D. (2014). *Stress test scenario: Sybil logic bomb cyber catastrophe.* Cambridge: Cambridge Risk Framew. Ser. Cent. Risk Stud. Univ. https://www.researchgate.net/publication/263262710_Stress_Test_Scenario_Sybil_Logic_Bomb_Cyber_Catastrophe
24. Sen, R., & Borle, S. (2015). Estimating the contextual risk of data breach: An empirical approach. *Journal of Management Information Systems*, *32*(2), 314-341. https://doi.org/10.1080/07421222.2015.1063315
25. Shrouf, F., & Miragliotta, G. (2015). Energy Management Based on Internet of Things: Practices and Framework for Adoption in Production Management. *Journal of Cleaner Production*, *100*, 235-246. https://doi.org/10.1016/j.jclepro.2015.03.055
26. Shu, X., Tian, K., Ciambrone, A., & Yao, D. (2017). Breaking the Target: An Analysis of Target Data Breach and Lessons Learned. *CoRR*, abs/1701.04940. https://arxiv.org/abs/1701.04940
27. Souri, A., & Hosseini, R. (2018). A state-of-the-art survey of malware detection approaches using data mining techniques. *Hum Centric Comput Inform Sciences*, *8*(3), 1-3. https://hcis-journal.springeropen.com/articles/10.1186/s13673-018-0125-x
28. Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications, 12,* 493-501. https://doi.org/10.1007/s12083-017-0630-0
29. Ucci, D., Aniello, L., & Baldoni, R. (2017) Survey on the usage of machine learning techniques for malware analysis. *Computers & Security, 81.* https://doi.org/10.1016/j.cose.2018.11.001
30. Zhu, J., Huang, H., & Zhang, D. (2017). «Big Tigers, Big Data»: Learning Social Reactions to China's Anticorruption Campaign through Online Feedback. *Public Administration Review, 79*(4), 500-513. https://doi.org/10.1111/puar.12866

Received 1.08.2020
Received in revised form 29.09.2020
Accepted 16.10.2020
Available online 21.11.2020