Journal of Namibian Studies, 35 S1 (2023): 1351-1362 ISSN: 2197-5523 (online)

Development Of The Intellectual System For Assessing Information Security Risks

Alibek Barlybayev [1], Gulnara Akimbekova [2]

[1] Astana International University, Astana, Kazakhstan
frank-ab@mail.ru

[2] Astana Medical University, Astana, Kazakhstan
akimbekova.g@amu.kz

Abstract

Rapid development of new digital technologies and their use, on the one hand, have opened up new opportunities for improving the efficiency of managing technological and business processes. On the other hand, it has led to significant increase in security threats, making businesses and organizations more vulnerable to cybercriminals. Soaring growth in the number of incidents of various types in recent years indicates insufficiency of traditional methods of ensuring information security (hereinafter IS). At present, consideration of the issue of IS from the point of view of risk management makes it possible to take cost-effective and effective enterprise security measures. In this paper, we are considering a project to create a system for assessing information security risks. In this paper escribed scientific novelty and significance of the project, also presented research methods used in the development of intelligent information security risk assessment systems.

Index Terms—information security risk assessment, fuzzy expert system, fuzzy multimodels, ontological model.

Introduction

The importance of IS in organizations can't be overestimated. It's important that organizations take the necessary steps to protect their priority information from unauthorized access, data leak, and other disruptive data security threats to business and client data.

Main problems of IS risk assessment:

- Great cost. Budgetary constraints can affect both the scope and scale of the assessment. Properly troubleshooting any cybersecurity issue requires a lot of time and expertise, which costs time and money.

- Incomprehensibility of goals scope. When dealing with large and complex environments, it can be difficult to determine overall goals and scope of assessment.
- Efficiency reporting. Reports must be accurate and complete. At the same time, they must do so in language that non-security personnel can understand. At minimum, baseline reports should be able to quickly identify areas that need special attention, as well as identify potentially standout units/departments that require further monitoring and attention.
- Large number of metrics and standards for assessing IS risks.

In this paper authors proposed several models to help various enterprises cope with IS problems [1]. This paper proposes a new hierarchical structured model for assessing IS risks using fuzzy logic. A new method for assessing software IS risks is also described using example of automated control systems. Proposed new risk assessment model was implemented programmatically in form of 15 fuzzy machines. In series of experiments, author carefully studied IS risk assessment of various software products. Model showed the strongest positive correlation with NIST-800-30, ISO/IEC TR-13335-3:1998, BS, expert opinion. However, other estimation methods have only one high correlation above 0.99 if proposed model is excluded from sample. t-test was in the range $-0.07758 \leq t_{fact} \geq 0.064225$. An increase in the sample reduced the variance of the mean and, therefore, increased sensitivity of test. With an increase in the number of degrees of freedom to 29, scatter of criterion narrowed significantly, i.e., became more powerful. And the mean value of the sample, although it didn't change, didn't fall into critical range. p-value remained large enough. Author made a statistical conclusion that the new proposed methodology is correct.

In this paper the audit is carried out according to methodology for conducting internal audit at enterprise, which is based on advanced standards and approaches to organization, management and security of IT infrastructure, such as Cobit, ISO-17799 [2].

Various methods are used to assess threats and vulnerabilities [3]:

- expert assessments;
- statistical data;
- consideration of factors affecting level of risk and vulnerability.

Mathematical apparatus of risk analysis is based on methods of probability theory related to probabilistic description of risks and uncertainties. In risk theory, the following types of mathematical models are distinguished: direct, inverse, and sensitivity research problems [4]. Direct risk assessment reports are based on previously known information to determine level of risk.

In risk analysis, the following class of mathematical models is widely used, taking into account uncertainty and differing in its description [4]:

- stochastic models;
- linguistic models;
- non-stochastic (game) models.

Risk assessment methods can be based on quantitative, qualitative and fuzzy logic depending on level of detail of available information and risks [5]. Numerical methods are often referred to as statistical methods: Monte-Carlo simulation [5], event tree analysis and regression [6], sensitivity analysis [6], expected annual loss [7], impact on risk [8], type of profitability and efficiency analysis [6] and others. On other hand, qualitative methods rely more on judgment than statistical methods: scenario analysis, fuzzy set theory [7] and others.

This paper discusses new method for assessing security risks in supervisory control and data acquisition (SCADA) networks using fuzzy logic [9]. Purpose of this paper is to improve the IS management analysis method by proposing a formalized approach, i.e., fuzzy analytical hierarchy process (AHP). This approach has been used to prioritize and select the most appropriate set of information security controls to meet the organization's IS requirements [10].

Scientific Novelty and Significance of the Project

One of main benefits of cybersecurity risk assessment is that it will help identify internal and external risks that are relevant to system. This is very important as it provides visibility into individual components of ICT security system and identifies which areas are weak and in need of improvement. This information will ultimately guide future security investments and provide guidance on how to improve.

Risk assessment enables to make more informed decisions by isolating each potential threat to vulnerability and calculating likelihood of risk occurring. Perhaps one of main reasons companies choose to assess their risks is to protect them from costly and damaging breaches. Risk treatment can help protect business from cyberattacks and improve protection of personal data.

It’s difficult (sometimes impossible) to make dozens of cybersecurity changes at same time for technical, operational and budgetary reasons. Assessment will help to justify which areas need better protection, prioritize which critical issues require attention in first place.

Detailed risk analysis will pinpoint which vulnerabilities are prioritized and why, as well as impact each could have on business if neglected. Once stakeholders and investors see what it will cost them if they don't make changes, they will be able to more favorably allocate risk treatment budget.

There is great social and economic demand for the provision of cybersecurity when using information and communication technologies: protection of personal data, protection of transmitted information, protection of electronic money accounts, protection of government systems like egov.kz, protection of nationally vital information resources and much more. Project results will make it possible to carry out a qualitative assessment of IS risks, reasonable choice of protection measures, to form a plan and budget for ensuring organization's security, thereby reducing business costs and government economic costs for occurrence of cybersecurity incidents. Loss of customer's and stakeholder's trust can be the most detrimental consequence of cybercrime, as the vast majority of people will not deal with a company that has been hacked, especially if it fails to protect its customers' data. This can lead to loss of business as well as depreciation of the brand.

Project results will allow scientists to take a fresh look at the issues of IS risk assessment. Structure of proposed new model of IS risk assessment criteria will be scientifically substantiated. Intelligent system will be developed that will allow determining a sufficient minimum investment and practical recommendations to reduce likelihood of IS incidents. All results of project will improve the level of IS protection and reduce costs in event of IS incidents. Expected results will influence methodology for creating IS policy in the production processes of organizations. They will also affect processes of accounting for IS in design, development and administration of software. The use of fuzzy models will allow optimizing (settings) of IS parameters when receiving measurement data from inputs and outputs from real systems. If knowledge generated by IS experts isn't accurate enough, then proposed new model will allow you to flexibly correct fuzzy models in system. If there is preliminary or partial knowledge about simulated IS system, they can be easily reflected in proposed new intelligent IS risk assessment system.

Innovativeness of the research is proved by project results:

- Dendogram and concept clusters used in IS risk assessment are new results. Previously, researchers have not conducted cluster analysis of IS risks. This result will be basis for building a new flexible IS risk assessment model. Also, this result will be proof of statistical correctness of new model structure.
- New flexible IS risk assessment model is an innovative component of project. Innovativeness of model is proved by its formalization through dendogram, clusters, thesaurus and ontological model. In fact, resulting ontology is a semantic mapping of the resulting new model.
- New intellectual IS risk assessment system is intelligent expert system that will add practical significance to project. Fuzzy multimodel expert system will be developed based on new flexible model obtained. To

improve model and prove its practical significance compared to other known models, statistical correlation experiments will be carried out.

Research Methods

The purpose of this study is to implement a flexible model for assessing IS risks. The issue of IS in information systems is very important for all types of organizations, as they usually allow you to manage all the main production business processes of the organization. A system with poor IS will eventually lead the company to huge financial losses. Given the lack of statistical data and the high uncertainty of the external environment, methods based on fuzzy expert systems using the experience and knowledge of employees can become the basis for sustainable economic development of an enterprise. The general purpose of fuzzy control systems is to model the thinking process of a person who draws conclusions for making any decision based on the available information about the control object. Situations of this kind are found in abundance both in everyday life and in the professional activities of people. The key to the successful use of such fuzzy-multiple methods in managing complex systems is the ability of systems to use all the main sources of information about the control object, which include: mathematical models; actual data of observations of the behavior of the object; knowledge of people – experts in the field under study. Indeed, all these sources can be used in a fuzzy control system, complementing each other. A mathematical model, if its construction is fundamentally possible and expedient, is the most important source of information that allows to replenish the knowledge base with the results of analytical research or simulation modeling. All this substantiates the problems of the proposed study. The processing of empirical data makes it possible to build an approximate model of the control object, as well as to clarify and adjust the parameters of the fuzzy control system. At the same time, based on the knowledge and experience of experts, a set of fuzzy rules is formed that reflect the patterns of behavior of the object under study. In cases where the development of a mathematical model is impossible due to the high complexity of the processes inherent in the control object, the advantages of methods based on fuzzy control are even stronger, since control is based not only on a certain model, but also intellectual control is implemented, which has in its basis of knowledge in all its variety of manifestations. Important advantages of fuzzy expert systems are non-linearity, the possibility of using inaccurate data, the convenience of obtaining and processing expert opinions.

The implementation of the Project will allow confirming the following hypotheses of the project:

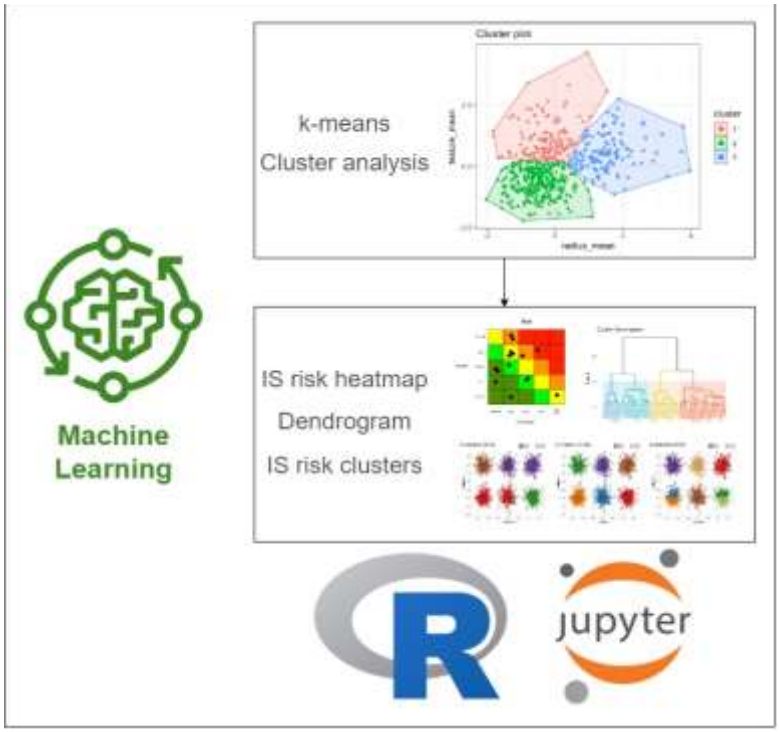
- The ability to assemble a flexible IS risk model that will meet the completeness criterion;

- The ability to build a production rule base that will meet the consistency criterion, despite the large flow of input variables;
- The possibility of assessing and auditing any information system according to various IS criteria;
- The possibility of obtaining expert conclusions on improving information security using fuzzy multimodels. The reporting conclusion will consist of the calculation of a sufficient minimum investment, practical recommendations for reducing the risks of information security and other parameters.

To proof of the above hypotheses, it is necessary to carry out scientific and technical work and research on the following methodologies:

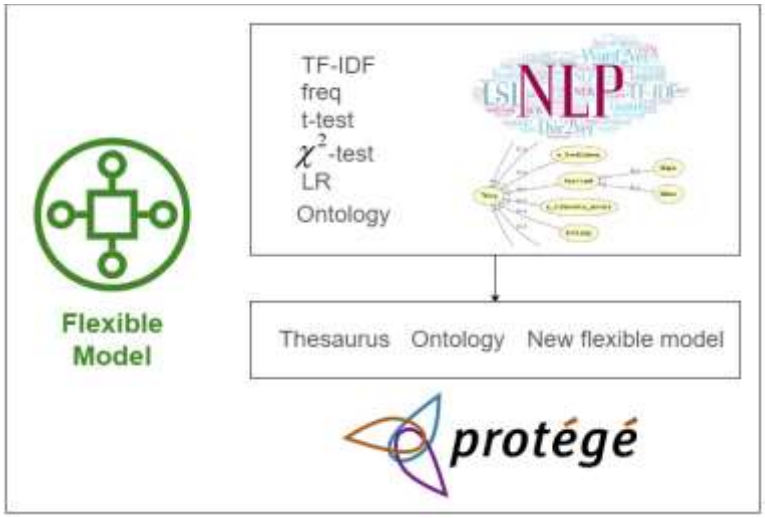
To solve the first task of the project, machine learning methods will be used. Figure 1 shows the architecture for solving the first task. Clustering will allow to determine which IS risks are similar to each other and potentially classify them. K-means clustering is needed to partition the data set into a set of k groups (i.e. k clusters), where k represents the number of groups pre-specified by the analyst [11]. The clustering of observations into groups requires some methods to calculate the distance or (dis)similarity between each pair of observations [12]. The result of this calculation will be a matrix of differences and distances. The distance measure determines how the similarity of two elements (x, y) is calculated and influences the shape of the clusters [13]. Objects will be classified into multiple groups (i.e. clusters) such that objects in the same cluster are as similar as possible (i.e. high intra-class similarity), while objects from different clusters are maximally different from each other (i.e. low inter-class similarity) [14]. In k-means clustering, each cluster is represented by its center (i.e. centroid) which corresponds to the average of the points assigned to the cluster. The main idea is to define clusters in such a way that the overall variation within a cluster is minimized. The standard algorithm is the Hartigan-Wong algorithm [15], which defines the total variation within a cluster as the sum of the squared distances, the Euclidean distances between elements, and the corresponding centroid.

Fig. 1 – Clustering to determine IS risks.

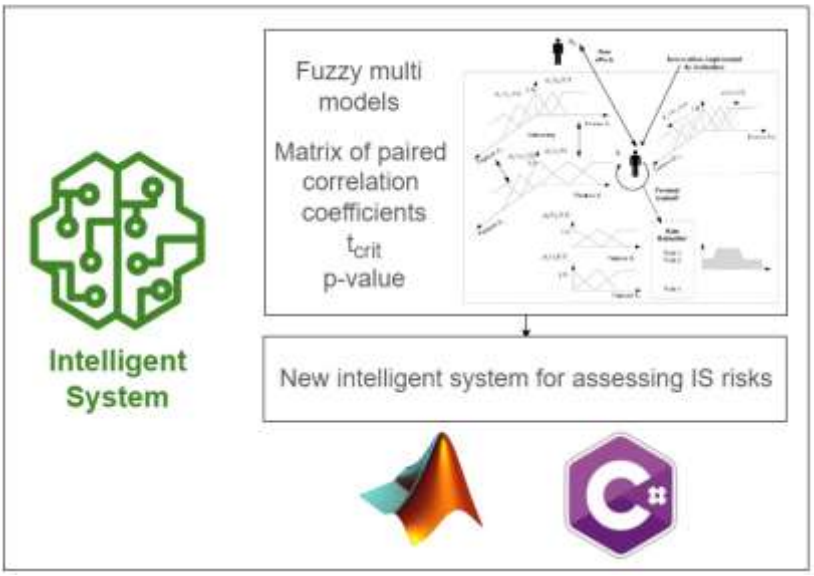


To solve the second task of the project, methods for creating thesauri of the subject area will be used. Figure 2 shows the architecture for solving the second task. Also, methods for constructing an ontological model of the subject area will be used. The thesaurus representation of information security risk assessment can be used for standardization and formalization, as well as to provide access to the knowledge of qualified and unskilled users who solve the problems of information security risk assessment. When constructing thesauri, nlp methods TF-IDF [16], collocation analysis [17], direct counting of the number of pairs (freq), t-test [18], χ^2-test [19], likelihood ratio (LR), on rules using templates will be used. TF-IDF will allow you to evaluate how relevant a word is in a set of documents that is part of a document collection or corpus. It does this by multiplying two measures: the number of times a word occurs in a document, and the reciprocal of the word's frequency in a document in a set of documents. The analysis of collocations is necessary for the selection of word combinations. Pair count (freq), t-test, χ^2-test, likelihood ratio (LR) – will be used to highlight two-word terms. An ontological representation of information security [20] risk assessment is necessary to represent knowledge in a convenient form for their processing by automated systems of semantic analysis. To build an ontology of IS risks, the framework for building knowledge bases Protégé will be used. The ontology will be formed by the Protégé-OWL modeling method, then exported to RDF Schema, XML Schema formats.

Fig. 2 – Creating thesauri of IS risks.



To solve the third task of the project, methods for creating fuzzy expert systems will be used. Figure 3 shows the architecture for solving the third task. A fuzzy multimodel for IS risk assessment according to Pedrycz [21-22] will be built. A multimodel is "a set of models M1, M2,...,Mc, provided with an appropriate mechanism for switching between models, or, if necessary, with a mechanism for aggregating the results provided by individual models". The operation of the switching mechanism is based on additional information about the system (its input parameters), and the aggregation mechanism is based on the confidence coefficients of individual models [23]. The use of fuzzy multimodels will allow solving the problem of "consistency" of the base of inference rules.



To verify the results of the research, a correlation test will be carried out to compare the developed new model with other known models for assessing IS risks. As a result, a matrix of paired correlation coefficients for all models will be obtained. To strengthen the confirmation of the reliability of the hypotheses, a statistical test will be carried out, the test sample will be increased by 5-10 times, the parameters will be calculated: t_{crit} - the critical value of the t-criterion, *p-value* - two-tailed could just happen by chance, two-sided Student's t-distribution [24-28].

To solve the scientific and engineering tasks of the project, the following software applications and technologies are supposed to be used: programming languages R [29], Python (Jupiter) [30] for solving the problems of information security risk clustering and nlp text processing; Protégé-OWL [31] for ontology design; Matlab for designing fuzzy expert systems [32]; C# for the development of the final software product of intellectual information security risk assessment [33].

Conclusion

The goal of the project is to create new flexible models and algorithms for assessing IS risks to improve the efficiency of IS management in an organization. Based on them, an intelligent system will be developed that will allow determining a sufficient minimum investment and practical recommendations to reduce likelihood of IS incidents.

Project objectives:

1) Research of IS risks using machine learning methods. Results: IS risk heatmap. Dendrogram of IS risks. IS risk clusters.
2) Development of flexible model of the most important concepts and relationships between them in subject area of IS risk assessment. Results: Thesaurus of IS risk assessment. Ontology of IS risk assessment. New flexible IS risk assessment model.
3) Development of intelligent IS risk assessment system. Results: New intelligent system for assessing IS risks. Description of existing IS risk assessment models. Correlation experiment results on resulting model.

Main benefits of cybersecurity risk assessment is that it will help identify internal and external risks that are relevant to system. This is very important as it provides visibility into the individual components of the ICT security system and identifies which areas are weak and in need of improvement. This information will ultimately guide future security investments and provide recommendations on how to improve system. Risk assessment enables us to make more informed decisions by isolating each potential threat to vulnerability and calculating likelihood of risk occurring. Perhaps one of main reasons companies choose to assess their risks is to protect them from costly and damaging breaches. Risk treatment can help protect business from cyberattacks and improve

protection of personal data. Detailed risk analysis will pinpoint which vulnerabilities are prioritized and why, as well as impact each could have on business if neglected. Once stakeholders and investors see what it will cost them if they don't make changes, they will be able to more favorably allocate risk treatment budget. Risk assessment offers solutions to protect information systems and allows you to reduce risks more intelligently. Project results allow scientists to take a fresh look at issues of IS risk assessment. Structure of proposed new model of IS risk assessment criteria will be scientifically substantiated. Intelligent system will be developed that will allow determining sufficient minimum investment and practical recommendations to reduce likelihood of IS incidents. Project's all results will improve level of IS protection and reduce costs in event of IS incidents. Project results can be commercialized as provision of services for assessment of IS, as well as audit with issuance of recommendations for improving IS of organizations.

Bibliography

1. Abdymanapov S.A., Muratbekov M., Altynbek S., Barlybayev A. Fuzzy Expert System of Information Security Risk Assessment on the Example of Analysis Learning Management Systems. IEEE Access. 2021, 9, pp. 156556–156565.
2. Mukatai G., Bekmanova G., Sharipbay A., Omarbekova A. The audit method of enterprise's Information security. ACM International Conference Proceeding Series. 2020, 3410761.
3. ISO/IEC 27005:2015 Information technology. Security techniques. Information security risk management.
4. Astakhov A.M. Iskusstvo upravleniya informatsionnymi riskami. M.: DMK Press. 2010, p. 312.
5. Bennett J.C., Bohoris G.A., Aspinwall E.M., Hall R.C. Risk analysis techniques and their application to software development. European Journal of Operational Research. 1996, vol. 95. pp. 467-475.
6. White D. Application of systems thinking to risk management: a review of the literature. Management Decision. 1995, №3(10). pp. 35-45.
7. Rainer R.K.J.R., Snyder C.A., Carr H.H. Risk analysis for information technology. Journal of Management Information Systems. 1991, №8(1). pp. 129-147.
8. Boehm B.W. Software Risk Management, IEEE Computer. Washington: Society Press. 1989, pp. 148-171.
9. Markovic-Petrovic J. D., Stojanovic M. D., Bostjancic Rakas S. V. A Fuzzy AHP Approach for Security Risk Assessment in SCADA Networks. Advances in Electrical and Computer Engineering. 2019, vol. 19, no. 3, pp. 69-74.
10. Tariq M. I., Ahmed S., Memon N. A., Tayyaba S., Ashraf M. W., Nazir M., Hussain A., Balas V. E., Balas M. M. Prioritization of information security controls through fuzzy AHP for cloud computing networks and wireless sensor networks. Sensors. 2020, vol. 20(5), 1310.

11. Huang T.-C., Huang C.-Y., Chen Y.-C. Real-Time DDoS Detection and Alleviation in Software-Defined In-Vehicle Networks. IEEE Sensors Letters, 2022, 6(9), 6003304.
12. Ma X., Keung J., Yang Z., Yu X., Li Y., Zhang H. CASMS: Combining clustering with attention semantic model for identifying security bug reports. Information and Software Technology. 2022, 147, 106906.
13. Yuan Y., Li Y. A Modified Hybrid Method Based on PSO, GA, and K-Means for Network Anomaly Detection. Mathematical Problems in Engineering. 2022, 5985426.
14. Alsanad A., Altuwaijri S. Advanced Persistent Threat Attack Detection using Clustering Algorithms. International Journal of Advanced Computer Science and Applications. 2022, 13(9), pp. 640-649.
15. Almanza-Ortega N.N., Pérez-Ortega J., Zavala-Díaz J.C., Solís-Romero J. Comparative Analysis of K-Means Variants Implemented in R. Computacion y Sistemas. 2022, 26(1), pp. 125-133.
16. Ghasiya P., Okamura K. A Hybrid Approach to Analyze Cybersecurity News Articles by Utilizing Information Extraction & Sentiment Analysis Methods. International Journal of Semantic Computing. 2022, 16(1), pp. 135-160.
17. Boontam P., Phoocharoensil S. Broaden Your Horizons: Distribution and Collocational Patterns of the English Synonyms Expand, Widen, and Broaden. International Journal of Communication and Linguistic Studies. 2022, 20(1), pp. 107-123.
18. García De Soto B., Turk Ž., MacIel A., Mantha B., Georgescu A., Sonkor M.S. Understanding the Significance of Cybersecurity in the Construction Industry: Survey Findings. Journal of Construction Engineering and Management. 2022, 148(9), 04022095.
19. Islam R., Refat R.U.D., Yerram S.M., Malik H. Graph-Based Intrusion Detection System for Controller Area Networks. IEEE Transactions on Intelligent Transportation Systems. 2022, 23(3), pp. 1727-1736.
20. Martins B.F., Serrano Gil L.J., Reyes Román J.F., Panach J.I., Pastor O., Hadad M., Rochwerger B. A framework for conceptual characterization of ontologies and its application in the cybersecurity domain. Software and Systems Modeling. 2022, 21(4), pp. 1437-1464.
21. Hu Q., Zhang L., Zhou Y., Pedrycz W. Large-Scale Multimodality Attribute Reduction with Multi-Kernel Fuzzy Rough Sets. IEEE Transactions on Fuzzy Systems. 2018, 26(1), 7805229, pp. 226-238.
22. Pedrycz, W. Fuzzy multimodels. IEEE Transactions on Fuzzy Systems. 1996, 4(2), pp. 139-148.
23. Wang, M., Wu, X., Tian, H., Lin L., He, M., Ding, L. Efficiency and Reliability Analysis of Self-Adaptive Two-Stage Fuzzy Control System in Complex Traffic Environment. Journal of Advanced Transportation. 2022, 6007485.
24. Razikin K., Soewito B. Cybersecurity decision support model to designing information technology security system based on risk analysis and cybersecurity framework. Egyptian Informatics Journal. 2022, 23(3), pp. 383-404.
25. Alqahtani M.A. Factors Affecting Cybersecurity Awareness among University Students. Applied Sciences. 2022, 12(5), 2589.
26. Alqahtani M.A. Cybersecurity Awareness Based on Software and E-mail Security with Statistical Analysis. Computational Intelligence and Neuroscience. 2022, 6775980.

27. Antunes M., Silva C., Marques F. An integrated cybernetic awareness strategy to assess cybersecurity attitudes and behaviours in school context. Applied Sciences. 2021, 11(23), 11269.
28. Yeng P.K., Fauzi M.A., Yang B. Assessing the effect of human factors in healthcare cyber security practice: An empirical study. ACM International Conference Proceeding Series. 2021, pp. 472-476.
29. Verhoeff J., Abeln S., Garcia-Vallejo J.J. INFLECT: an R-package for cytometry cluster evaluation using marker modality. BMC Bioinformatics. 2022, 23(1), 487.
30. Vijayalakshmi J., Ramaraj E. A Hadoop-big data analytic model to predict and classify chronic kidney diseases using improved fractional rough fuzzy K-means clustering and extreme gradient boost rat swarm optimizer. Concurrency and Computation: Practice and Experience. 2022, 34(28), e7354.
31. Alazzam H., Abualghanam O., Al-Zoubi Q.M., Alsmady A., Alhenawi E. A New Network Digital Forensics Approach for Internet of Things Environment Based on Binary Owl Optimizer. Cybernetics and Information Technologies. 2022, 22(3), pp. 146-160.
32. Akhmetov B., Lakhno V., Malyukov V., Akhmetov B., Yagaliyeva B., Lakhno M., Yakiyayeva G. A Model for Managing the Procedure of Continuous Mutual Financial Investment in Cybersecurity for the Case with Fuzzy Information. Lecture Notes on Data Engineering and Communications Technologies. 2022, 93, pp. 539-553.
33. Ullah F., Ali Babar M., Aleti A. Design and evaluation of adaptive system for big data cyber security analytics. Expert Systems with Applications. 2022, 207, 117948.