Determining Key Risks for Modern Distributed Information Systems

Dmytro Palko, Hrygorii Hnatienko, Tetiana Babenko and Andrii Bigdan

Taras Shevchenko National University of Kyiv 64/13, Volodymyrska Street, Kyiv, 01601, Ukraine

Abstract

This work aims to study the problem of identifying and assessing information security risks in complex, distributed, and scalable information systems, as well as building a profile of key risk factors that can cause potential information security incidents in the physical and functional allocation of resources. As part of this work, a study was carried out of the main information security risks that can be identified at the time of creating and operating a typical distributed information system designed to support information processes and provide information services. The result of the study is the ranking of major risk factors according to their importance and frequency in practice, as well as highlighting the most significant security controls. The data for analysis was compiled based on the results of interviews and questionnaires of information security specialists with different training levels and different focuses in their activities within this knowledge area. The paper presents summarized information on classical approaches to information security risks modeling based on quantitative, qualitative, and hybrid analysis, as well as the latest methodologies based on solving the problems of intelligent classification and analysis of data on risk factors in the system distribution, and in operation with large data sets.

Keywords[1]

Information security risk, distributed information systems, risk factors, security controls, risk control techniques, risk modeling, risk management, risk assessment, intelligent risk assessment models

1. Introduction

Today, information security management plays a key role in the life processes of almost any organization that uses modern technologies for collecting, processing, and storing information. This process is based on the regular assessment of information risks, which allows you to timely identify new threats and vulnerabilities, implement appropriate measures to neutralize them, and continuously monitor the state of information security of the system, considering the previous experience and new factors.

To prepare for potential attacks and possible problems of this nature, as well as to prevent disruptions to business processes and operations, damage to reputation, or loss of data, organizations must constantly assess their risk profile, make recommended corrections, and actively improve their security system. Threat analysis and risk management are the cornerstones of any security policy. Cybersecurity risks should be considered as a key factor in the strategic planning of business processes. That is why it is the responsibility of each company to develop a risk assessment methodology that best suits the organization's priorities and business goals.

The importance of risk management as a process in modern reality is undeniable. The modeling and forecasting information security risks task has been and remains a significant and priority. This issue is especially relevant in the context of the widespread of complex multi-component information systems

II International Scientific Symposium «Intelligent Solutions» IntSol-2021, September 28–30, 2021, Kyiv-Uzhhorod, Ukraine
EMAIL: palko.dmytro@gmail.com (A. 1); g.gna5@ukr.net (A. 2); babenkot@ua.fm (A. 3); abigdan@gmail.com (A. 4)
ORCID: 0000-0002-2886-1975 (A. 1); 0000-0002-0465-5018 (A. 2); 0000-0003-1184-9483 (A. 3); 0000-0002-2940-6085 (A. 4)

that have a distributed nature and contain a large number of nodes. The total number of computer systems, active network equipment, and peripherals installed in any infrastructure is growing at a phenomenal rate. The relative simplicity of networking is a compelling reason to interconnect computer systems, share functions, and share resources. This approach allows better use of a computing power vast set that is currently available, but on the other hand, raises some issues primarily related to the complexity of security and potential risk management. The transition to complex, large-scale, and structurally complex information systems increases the likelihood of unforeseen and unplanned events affecting the performance and operability of the system as a whole.

2. Literature Review and Problem Statement

2.1. Distributed Information Systems overview

The widespread introduction of distributed information systems (DIS) today is representative of almost all areas of human activity, where they are entrusted with solving more and more important tasks. The efficiency of decision-making and the efficiency of the functioning of many economic, social, political, and military structures depend on the quality of DIS functioning.

Distributed information systems are complex technical systems consisting of many structural elements, functionally combined to provide one or more types of information processes and the provision of information services. Such systems typically operate under random factors, the presence of negative influences of various natures, active interaction with the external environment, and the high cost of impacts of possible violations or malfunctions. All this causes many problems related primarily to information security. Managing the cybersecurity risk assessment in distributed systems involves solving a set of problems related to functional distribution and hierarchy, a high degree of resources parallelization, and a near-complete lack of centralized management.

On the way, there are difficulties to the analysis of heterogeneous data, the need to reconcile information obtained from different sources, the variability of distributed metrics, which requires a wide arsenal of tools for analytical processing and intelligent data processing of different nature, the problem of incomplete information about the components of a distributed system and the complexity of integrated multifactor analysis in general. There is a need, on the one hand, for a set of methods and tools that can eliminate these obstacles, and on the other hand, for a new approach to organizing research on information security risks in a distributed environment and performing comprehensive analytical processing of distributed data of various natures. Therefore, the implementation of a new approach to managing risk assessment in DIS involves the introduction of a comprehensive solution that integrates data obtained from different sources and a wide range of tools for their analysis [1].

Several international organizations and leading universities are engaged in research on this issue. The key standards in this area that need to be relied on are ISO / IEC 27001: 2013, NIST SP 800-30, and BS 7799-3: 2017. However, despite significant achievements, there is currently no single system vision of all aspects of the problem, the nature, and features of research tools, and its place in the process of multifactor risk analysis of a distributed system, considering the entire complex of interrelations and mutual influence of the processes associated with it. The different degree of depth of elaboration of certain aspects of this problem has led to the need for effective models and methods of reconciliation and analytical processing of heterogeneous data for rapid analysis of the current state of information security of a distributed system.

2.2. Risk Management Process in Distributed Information Systems

Information security risk assessment is an extremely important part of a company's data protection strategy. It is conducted out to support decision-making and immediate response to identified threats (risk response).

Information security risk analysis allows you to determine the necessary and sufficient set of information security tools, regulatory and organizational mechanisms to reduce information security risks, allowing to ensure the process of building the most effective information security management system architecture for a given organization [2].

Risk management is an iterative process of identifying, quantifying, analyzing, and managing the risks faced by an organization [3]. Risk management is designed to ensure a stable operation of the information system and minimize possible losses in the event of information security threats. As an integral part of management practice, risk management should be carried out regularly to support organizational improvements, improve existing security tools and mechanisms, improve efficiency and make management decisions [4]. The main risks are those risks that have a high likelihood of occurrence and, if implemented, provide the possibility of a significant impact on operational performance, achievement of the goals and objectives of the project, or may damage reputation [5].

In terms of systemic distribution, risk management should provide for a complex nature, and consider the risk assessment for each asset or subsystem.

The specificity of the architecture of distributed information systems involves the analysis of data, largely differentiated in their structure, and the use of all available tools of assessment methods (both quantitative and qualitative) that characterize various components of the studied environment. The poor structure of the tasks of such research resulted from the lack of formal models and obtaining objective measurements results together with subjective expert assessments.

Thus, the risk management process in distributed information systems is a sophisticated and rather integrated task.

The study of risk factors in a distributed environment deserves special attention.

According to ISACA's annual STATE OF ENTERPRISE RISK MANAGEMENT 2020 survey [6], the biggest challenges in corporate risk are factors related to the emergence of new threats, changes/advances in technology development, as well as weak human resources and lack of necessary skills and experience of specialists and existing cybersecurity teams (Figure 1).

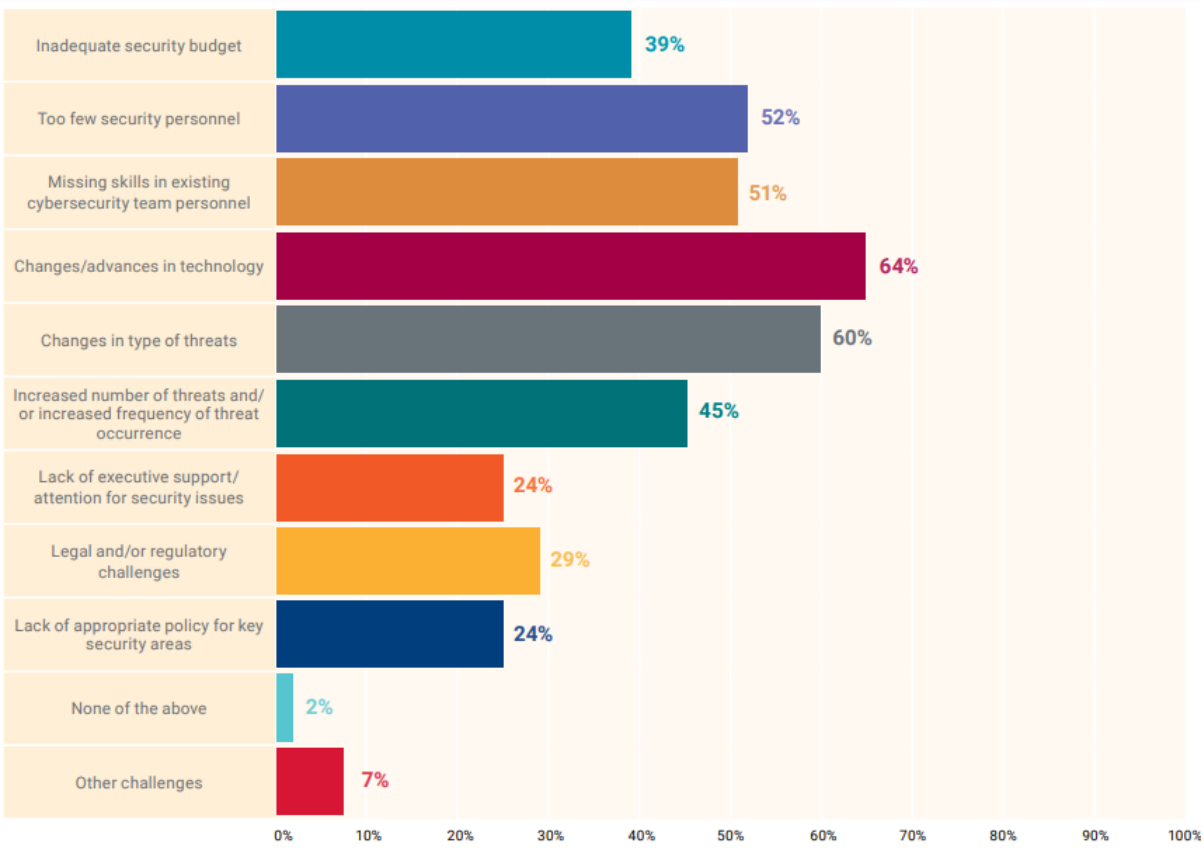

Figure 1: Cybersecurity challenges today

On the other hand, according to this study, the most frequently used control to prevent/mitigate potential security concerns is to raise awareness and conduct training on cybersecurity among staff (Figure 2).

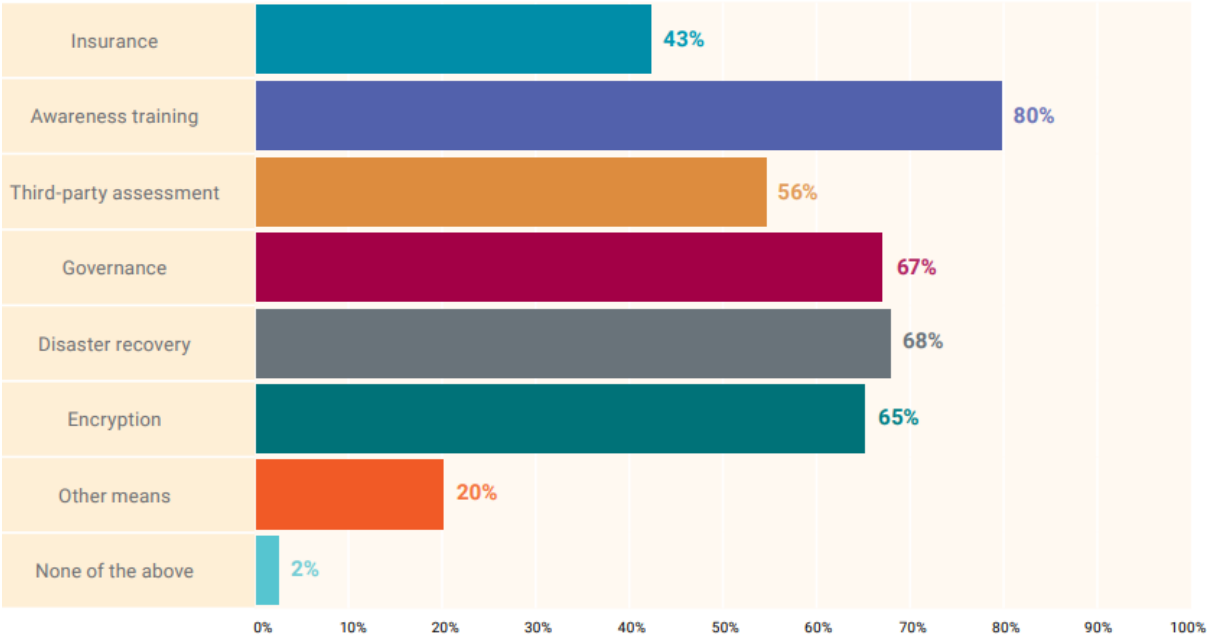

Figure 2: Top mitigation controls

Eighty percent of enterprise respondents provide awareness-raising training, 68 percent use disaster/incident recovery strategies, and 67 percent use general information security control and management. Less than half of the responding businesses use insurance as a mitigation control; at the same time, the largest supporters of this approach are companies in North America and Africa [6].

2.3. Main Approaches to Risk Assessment

There are many different methods for analyzing information risks for distributed systems. Their main differences are the approaches and the scales being used for assessing the risk level: quantitative or qualitative.

Conventionally, among the methods of risk assessment, the following three groups can be distinguished:

1. Statistical methods
2. Methods of expert assessments
3. Modeling methods

2.3.1. Statistical Risk Assessment Methods

To assess the information security risks, a qualitative, quantitative, or combined approach can be used.

In quantitative methods, the risk is assessed in the form of numerical values. Accumulated statistical information on incidents and violations, as well as meta-information about the current state and configuration of the node components of the distributed system, are usually used as input data for the assessment. However, the frequent lack of sufficient statistics leads to a decrease in the adequacy of the assessment results. Other limitations are complexity, high labor intensity, and long execution time, especially in the terms of the analysis of distributed systems. The advantages of the quantitative approach include the accuracy of risk assessment, clarity of results, and the ability to compare the risk value, expressed in financial equivalent with the investment amount required to respond to this risk.

Qualitative methods are more common, but they use too simplified scales, usually containing three levels of risk assessment (high, medium, low). The assessment is based on expert surveys, and promising intellectual methods are still insufficiently applied. Other disadvantages are the lack of visibility and complexity of using the results of risk analysis for economic justification and assessing

the feasibility of investing in risk response measures. The advantage of a qualitative approach is its simplicity and minimization of the time and labor costs for conducting a risk assessment [7].

The combined approach involves a combination of both methods to apply the benefits of each.

According to "The Marsh Microsoft 2019 Global Cyber Risk Perception Survey" (September 2019), the popularity of a quantitative approach to assessing information security risks has increased significantly compared to 2017, but it remains low (Figure 3) [8].

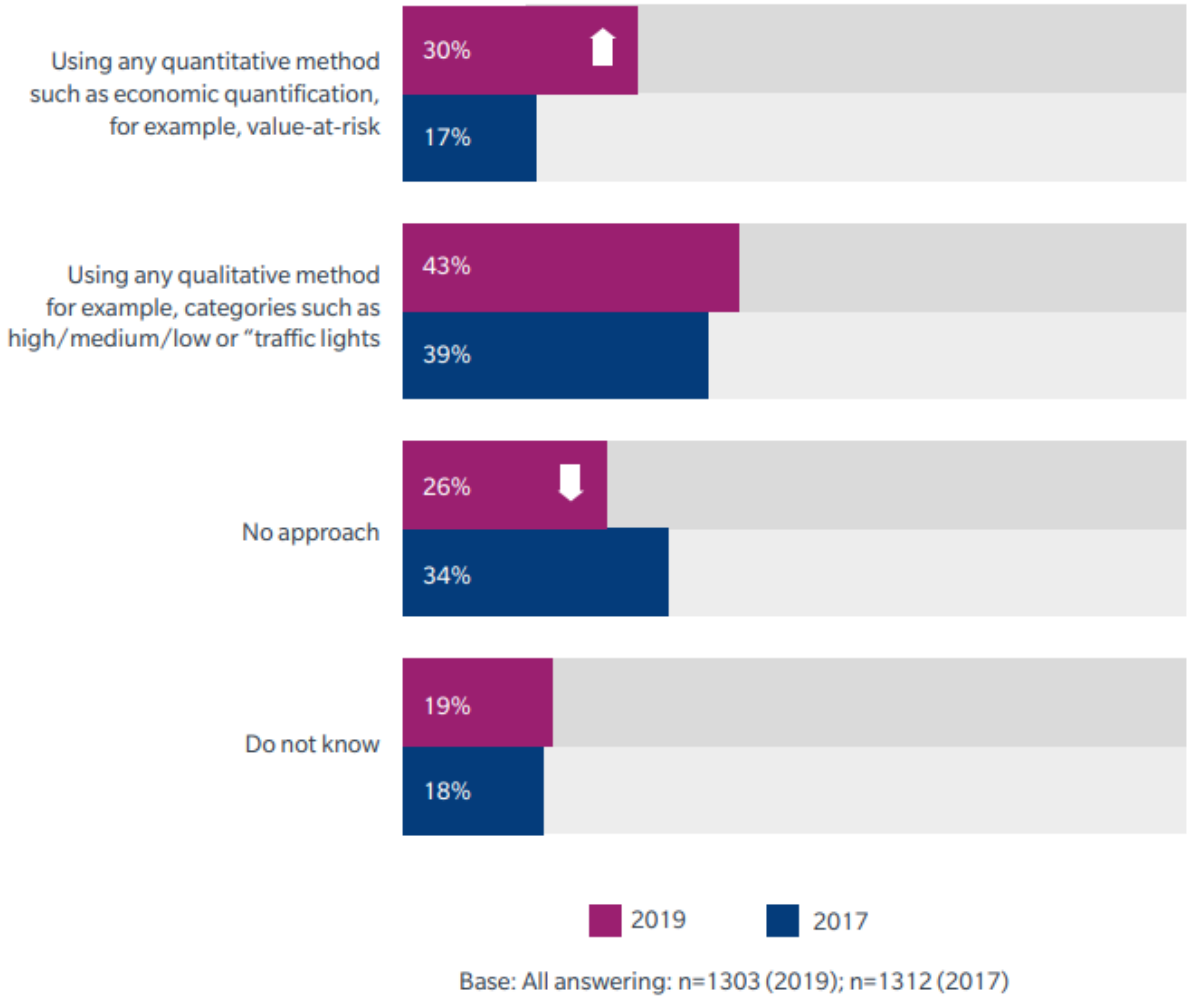

Figure 3: Approaches to assessing information security risks following "The Marsh Microsoft 2019 Global Cyber Risk Perception Survey"

Thus, today most companies use a qualitative scale to assess information security risks.

2.3.2. Statistical Risk Assessment Methods

If it is impossible to use statistical methods for analyzing the risks of a distributed system (lack of information on risk factors, insufficient data sampling size, complexity and sophistication of infrastructure, etc.), you should refer to expert assessment methods. The essence of the method of expert assessments is to conduct an expert analysis of the problem using qualitative and quantitative assessment of hypotheses and further processing of the results. This method is simple and accessible for practical application but requires a significant level of competence and extensive practical experience from the expert.

When using expert methods, the risk level is assessed based on the analysis of the probability of an adverse event occurring by studying and assessing the factors affecting it. Thus, the practical application

of this method is to establish a list of factors that determine a particular type of risk, as well as to determine the relationship between the nature of the factor and the risk level that this factor causes.

For the objectivity and impartiality of the results, the work on identifying and assessing information security risks should be carried out by special experts or relevant expert groups who have the necessary experience and training on this matter issue [9].

2.3.3. Modeling Methods

The most effective methods for analyzing information security risks in distributed systems are modeling methods [10], among which there are neural networks that can identify and adequately assess information security risk relying on data mining tools. The need to extract unknown, non-trivial, practical, and useful knowledge from the "raw" metadata about the operation of a distributed system, which can be interpreted in a certain way and used to make decisions about the risk level, gives this problem a non-trivial interpretation.

Artificial Intelligence (AI) is not a new concept, but only in recent years, various companies have begun to explore and understand its full potential. Intelligent systems play an increasingly important role in network management. Most research in intrusion detection and risk assessment systems heavily relies on AI techniques to design, implement, and improve security monitoring systems.

Recent malware updates and improvements in cyberattacks are difficult to detect with traditional cybersecurity techniques. An important advantage of neural networks is their ability to "learn" the characteristics of the input data and identify elements that are not similar to those previously observed in the system. New AI algorithms use machine learning to quickly adapt and analyze new data, improve results and identify new vectors of risk implementation.

Most modern methods of attack detection and risk assessment leverage some form of rule-based analysis or a statistical approach. The analysis relies on a set of predefined rules created by the administrator or by the security system itself.

Unlike expert systems, which can give the user a definite answer about the compliance of the considered characteristics embedded in the knowledge base rules, the neural network analyzes the information and provides an opportunity to assess, reconcile the data with the characteristics it is trained to recognize [11].

Thus, forecasting and modeling the level of risk, coordination, and intelligent processing of various nature data about risk factors and creating based on their analysis a comprehensive approach to risk assessment in distributed information systems is a priority research area today.

3. The Research Methodology

The purpose of this study is to highlight the key risk factors inherent in modern distributed information systems, analyze the most significant security controls for development based on their recommendations to eliminate potential threats.

3.1. The Data Collection Process for Analysis

A questionnaire method was used to collect a sample of test data for analysis. The respondents were several dozen information security engineers of various training levels, penetration testing and auditing specialists, and leading specialists in the field of information security project management. All interviewees were selected selectively and have experience in providing security for information system infrastructures of various sizes and scales.

The study involved two data collection processes for analysis. The first is a pilot survey to test the research instruments and adjust the questions, the second is a mass survey of the target group using the final version of the questionnaires.

The pilot study was performed before the main questionnaire and aimed to test whether the proposed model of the questionnaire is suitable for the analysis of the final metrics.

23 specialists (Table 1) were involved in the main survey. Age of survey participants from 24 to 47 years, with an average of 34.2.

3.2. Questionnaire Development

The development of questionnaires considers the most common risk factors that are common to most modern distributed infrastructures. The questionnaire included 40 questions on the main risk factors and 14 questions on the practice of applying security controls in real projects. These indicators were identified at the stage of analysis of literature sources in this subject area and during the pilot survey. The respondents were asked to answer these questions anonymously and subjectively, relying on their own experience and the real practice of working with distributed information systems.

IBM SPSS Statistics software toolkit was used for data analysis and modeling as it is widely used for statistical analysis by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others. The research findings are analyzed and discussed in the following sections.

Table 1

The Survey Participants Classification

1	2	3	4
№	Categories	Number of respondents	Percentage (%)
1.	Gender		
1.1.	Female	6	26.08
1.2.	Male	17	73.91
	Total	**23**	**100.00**
2.	Position		
2.1.	Information Security Engineer	6	26.08
2.2.	Information Security Auditor	2	8.69
2.3.	Penetration Tester	5	21.73
2.4.	Malware Analysts	2	8.69
2.5.	Infrastructure Engineer[2]	5	21.73
2.6.	Project manager	3	13.04
	Total	**23**	**100.00**

The share of female respondents among the interviewers is only 26 percent.

3.3. Research Criteria and Analysis of the Test Sample

The respondents were asked various questions that used scales from 1 to 5. To increase efficiency and narrow the gradation of possible results for assessing risk factors at work, a 5-point scale was chosen, in which the indicator “does not matter” is equal to one, and “extremely important” is equal to five. Likewise, there are five categories for assessing security controls so that the “never” indicator is one and the “always” indicator is five. Thus, all questions about risk factors in distributed systems were measured on a five-point Likert scale from “nonsignificant” to “most important”, and all security controls – from "never" to “always”. The Likert scale is quite easy to build, it provides relative reliability even with a small number of judgments, and the data obtained is easy to process. The selection of judgments for the scale was carried out based on an analysis of literary sources in a given subject area and during pilot research by the method of selection from an initial list of judgments with the most discriminatory ability to the measured attitude. For this purpose, an initial list of statements was created (Table 2) that were offered to respondents from a group representative of the target audience (participants in the pilot study).

[2] With a background in the field of cybersecurity

Table 2
Risk Factors and Security Controls Measures Scale

Scale	Risk factors	Security controls
1.	Unimportant	Never
2.	Slightly Important	Seldom
3.	Important	Sometimes
4.	Very Important	Often
5.	Critical	Always

When working with the scale, the respondents rated the degree of their agreement or disagreement with each of the proposed judgments, from "completely agree" to "completely disagree".

3.4. Key Risk Factors

The study demonstrates 40 main risk factors in modern distributed information systems (Appendix 1, Table A1), labeled from Factor_1 to Factor_40, which are quite common in the relevant literature, are often found in practice, and are widely used by researchers and experts in cybersecurity when studying risk factors and conducting risk management measures. These factors should be identified in the process of assessing and managing information security risks and monitored in the future.

Separately, the risks caused by a human factor should be highlighted. They include not only employee mistakes, but also intentional actions that lead to violating information confidentiality.

Referring to the NIST SP 800-37 Risk Management Framework, should not forget about such categories shown in Table 3.

Table 3
Risk Categories

№	Category
1.	Financial risks
2.	Legal risks
3.	Business risks
4.	Political risks
5.	Software risks
6.	Risks of non-compliance with legislation
7.	Security and confidentiality risks
8.	Project risks
9.	Reputational risks
10.	Risks of life safety
11.	Risks of strategic planning

They were not considered in this study, however, constitute an important part of any risk management process [12].

3.5. Key Security Controls and Risk Management Measures

As a result of the analysis of the above statistical data, the expert group proposed possible categories of actions to minimize information security risks, including organizational and legal protection of information, engineering, hardware and software protection, cryptographic mechanisms for protecting information [13], as well as institutional arrangements and physical protection measures.

As effective controls to ensure the security of distributed systems by trained full-time specialists or with the help of information security outsourcing, the following solutions (both separately and in aggregate) can be implemented, shown in Table 4. The ISO 27001 standard and its Appendix A are

important tools for information security management [14] and it was a ground for developing questionnaires on possible control and risk management measures. It contains a list of security measures that must be applied to improve information security and consists of 114 security controls, divided into 14 chapters. Not all of these controls are mandatory for implementation – the company can choose on its own, it considers the controls applicable in the given circumstances and depending on the business direction, infrastructure state, or the existing profile of external threats, and then implement them (usually at least 90 % controls). A more detailed description of each control in Appendix A with an explanation of how it should be applied is presented in the ISO 27002 standard. However, the latter does not provide any explanations and tips on how to choose control in a given situation, which controls to implement, how to measure them and how to distribute duties [15].

Table 4

Security Systems for Implementing Technical Security Controls

1	2
№	Protection system
1.	Backup and restore systems
2.	Protection system against unauthorized access
3.	Network shielding systems
4.	Protection systems against attacks at the application level (WAF)
5.	Incident and event management systems (SIEM)
6.	Identity and Access Management systems (IAM)
7.	Security and confidentiality risks
8.	Management systems for compliance with information security requirements (Compliance Management)
9.	Data leak prevention systems (DLP)
10.	Information right management systems (IRM)
11.	Solutions for network security
12.	Anti-virus protection systems
13.	E-mail protection systems
14.	Content filtering systems for web traffic
15.	Access control systems to peripheral devices and applications
16.	Systems for monitoring the integrity of software environments
17.	Cryptographic protection for stored information

Separately, it should be noted the international standard ISO/IEC 27005: 2018 “Information technology – Security techniques – Information security risk management”, which contains recommendations for information security risk management. This document supports the general concepts defined in ISO/IEC 27001 and is intended to guide the implementation of information security measures based on a risk-based approach [16]. In the study, it was proposed to evaluate 14 main groups (Appendix 1, Table 6) of information security controls of modern distributed information systems in terms of frequency and effectiveness of their use, labeled from Control_1 to Control_14.

Thus, the proposed options cover the entire range of the most common risk management mechanisms and measures used in modern distributed systems.

4. Results and Discussion

4.1. The Importance of Risk Factors in Lifecycle of Modern Distributed Information Systems

Table 7 (Appendix 1) shows that nearly all respondents ranked factors related to lack of cybersecurity policy, lack of protection mechanisms against network attacks, violations of

authentication and session management, violations of access control, and use of components with known vulnerabilities as the most important.

The uncorrected sample standard deviation *S* is calculated (1) for four groups of factors, each of which contains 10 of them

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^2}, \quad (1)$$

where $\{x_1, x_2, \ldots, x_n\}$ are the mean values of the sample items, $n = 10$ – the size of the sample (number of factors in each group: Factor_1 – Factor_10, Factor_11 – Factor_20, Factor_21 – Factor_30, Factor_31 – Factor40), and $\tilde{x}$ – the mean value of this assessment (2)

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \quad (2)$$

The vast majority of technological factors play a key role in determining the final risk level. Analysis and summarization of the survey responses gave the following rating of the importance of the listed risks (in order of importance): Factor_20, Factor_6, Factor_10, Factor_11, Factor_12, Factor_14, Factor_8, Factor_9, Factor_4, Factor_3, Factor_13, Factor_17, Factor_18, Factor_1, Factor_15, Factor_15 , Factor_5, Factor_7, Factor_16.

Among organizational factors, the risks associated with the lack of cybersecurity and anti-virus protection policies are the most important. The ranking of the importance of risks in this category (in order of importance): Factor_21, Factor_27, Factor_30, Factor_28, Factor_29, Factor_25, Factor_23, Factor_24, Factor_22, Factor_26.

In addition, all respondents noted that the risk of abuse of privileges is the highest risk factor and very important among the factors associated with the human factor. Risk severity rating for this category (in order of importance): Factor_33, Factor_40, Factor_35, Factor_37, Factor_38, Factor_36, Factor_34, Factor_31, Factor_32, Factor_39.

In summary, the categories of risk factors can be ranked in order of importance and criticality as follows: logical, physical, human factors, and organizational factors.

Table 5 illustrates a list of the top 10 key risk factors for distributed information systems based on a survey of experienced cybersecurity managers and engineers.

Table 5

Top 10 Risk Factors for Distributed Information Systems

№	N	Mean	Std. Deviation	% percent
Factor_21	23	4.391304	0.656376	87.8260
Factor_20	23	4.304348	0.634950	86.0869
Factor_6	23	4.217391	0.735868	84.3478
Factor_10	23	4.173913	0.650327	83.4782
Factor_11	23	4.130435	0.625543	82.6087
Factor_12	23	4.086957	0.668312	81.7391
Factor_33	23	4.000000	0.738549	80
Factor_27	23	3.956522	0.824525	79.1304
Factor_14	23	3.913043	0.792754	78.2608
Factor_8	23	3.869565	0.694416	77.3913

4.2. Frequency of Controls Occurrence

Table 6 shows the mean and standard deviation for each group of security controls. The results of this study show that most security controls are used frequently and are important mechanisms to prevent and minimize potential risks.

4.3. Construct Validity (Risk Factors Correlation)

The next step was to test the hypothesis about relationships between key risk factors using correlation coefficients.

The correlation coefficient is a statistical indicator of the probability of a relationship between two variables, measured on a quantitative scale, which allows you to answer the question of the degree and direction of the relationship between the values of these variables.

To choose the right method of correlation research, it is necessary to answer the question of whether the studied factors are normally distributed. Frequency histograms for key risk factors are presented in Appendix 2. An example of a histogram for factor Fact_21 is shown in the Figure 4.

Table 6

The Mean Score for Each Control Factor

1 №	2 N	3 Mean	4 Std. Deviation	5 % percent
Control_1	23	4.260870	0.619192	85.2174
Control_2	23	3.391304	0.940944	67.82608
Control_3	23	2.347826	1.070628	46.95652
Control_4	23	3.782609	0.735868	75.65218
Control_5	23	4.304348	0.634950	86.08696
Control_6	23	4.478261	0.593109	89.56522
Control_7	23	4.478261	0.665348	89.56522
Control_8	23	4.260870	0.619192	85.21740
Control_9	23	4.130435	0.625543	82.60870
Control_10	23	3.000000	1.044466	60.00000
Control_11	23	2.086957	0.900154	41.73914
Control_12	23	2.130435	0.757049	42.60870
Control_13	23	2.478261	0.845822	49.56522
Control_14	23	2.782609	0.951388	55.65218

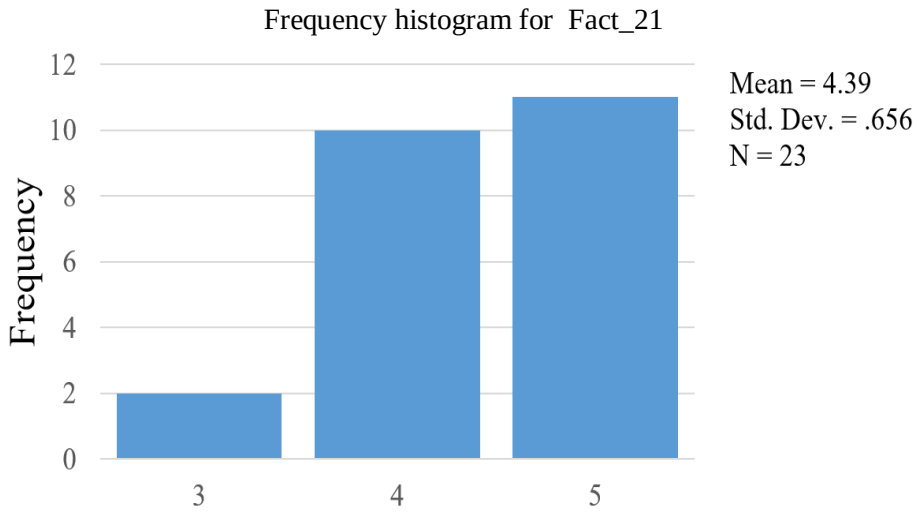

Figure 4: Example of frequency histogram for risk factor Fact_21

There are two hypotheses (3) for the test:

- Null hypothesis (H_0): the data comes from the specified distribution.
- Alternate Hypothesis (H_1): at least one value does not match the specified distribution.

That is,

$$H_0: P = P_0, H_1: P \neq P_0, \quad (3)$$

where P is the distribution of our sample and P_0 is a normal distribution.

Even though the plotted frequency histograms at first glance are quite symmetric and are well described by the parabolic curve for both tests, significance values less than .05 which means that the data do not have a normal distribution (Figure 5). So, the null hypothesis that the data is normally distributed was rejected.

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Fact_21	.301	23	<.001	.760	23	<.001
Fact_20	.293	23	<.001	.771	23	<.001
Fact_6	.248	23	<.001	.798	23	<.001
Fact_10	.301	23	<.001	.788	23	<.001
Fact_11	.322	23	<.001	.778	23	<.001
Fact_12	.291	23	<.001	.798	23	<.001
Fact_33	.239	23	.001	.815	23	<.001
Fact_27	.260	23	<.001	.857	23	.004
Fact_14	.223	23	.004	.807	23	<.001
Fact_8	.270	23	<.001	.804	23	<.001

a. Lilliefors Significance Correction

Figure 5: Test of Normality for Key Risk Factors

Since the volume of the studied sample is small (n<30), all factors are quantitative and the distribution of their values is not normal, it is decided to choose the rank correlation coefficient r-Spearman (4).

Deciding on the type of correlation when interpreting the results, it is important to remember and keep in mind that linear correlations are more accurate than rank correlations. Ranking of values when using r-Spearman naturally reduces the degree of individual variability of the measured indicator.

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \quad (4)$$

where $n = 10$ – number of factors, d_i is the difference between the two ranks of each assessment.

To assess the feasibility of using the above-described research tools, the correlation coefficients were calculated for key risk factors of modern distributed information systems.

The interpretation of the correlation coefficient is based on the level of the bond strength:

- $0.70 < r \leq 1.00$ – strong positive connection,
- $0.30 < r \leq 0.69$ – moderate positive connection,
- $0.01 < r \leq 0.29$ – weak positive connection,
- $-0.01 > r \geq -0.29$ – weak negative connection,
- $-0.30 > r \geq -0.69$ – moderate negative connection,
- $-0.70 > r \geq -1.00$ – strong negative connection.

The interpretation of the significance level (p-value) of the correlation coefficient is carried out in the same way as it was done for parametric and nonparametric criteria:

- if the p-value ≤ 0.05, the relationship between variables is statistically significant;
- if the p-value > 0.05, the relationship between variables is statistically nonsignificant.

Also, when interpreting the p-value of the correlation coefficient, it is important not only the fact of significance but also its level. Traditionally, the p-value of correlation is differentiated into three levels:

- $.01 < p \leq .05$ – low statistical significance (one star – *),
- $.001 < p \leq .01$ – the average strength of statistical significance (two stars – **),
- $p \leq .001$ – high statistical significance (three stars – ***).

Table 10 (Appendix 1) illustrates the relationship between key factors.

The correlation analysis revealed a moderate negative relationship of medium statistical significance between factors Factor_20 and Factor_8 – *r*-Spearman =-0.528 at $p \leq .01$, as well as a moderate negative relationship of low statistical significance between factors Factor_14 and Factor_8 – *r*-Spearman =-0.415 at $p \leq .05$.

Analyzing the results of correlation analysis, we can conclude that among the studied risk factors there is a moderate positive relationship of low statistical significance for the correlation of variables Factor_10 and Factor_11 – *r*-Spearman = 0.423 at $p \leq .05$.

Thus, the obtained results indicate that risk factors are often interrelated and have complex impacts, and therefore require a comprehensive and multidisciplinary analysis, considering all possible factors and conditions.

5. Conclusions

Thus, this paper investigates the problem of identifying and assessing information security risks in complex, distributed, and large-scale information systems, and also builds a profile of key risk factors that can cause potential information security incidents in the physical and functional allocation of resources. The study examines the main risks of information security that can be identified during the construction and operation of a typical distributed information system designed to provide one or more types of information processes and provisioning information services. The result was a ranking of the main risk factors according to their importance and frequency in practice, as well as highlighting the most significant security controls.

Thus, the results of the study show that all risk factors in the life cycle of a modern distributed system are very important and require detailed analysis and consideration when building a profile of potential threats and assessing information security risks. The importance rating of the risk factors categories by nature can be given as follows (in order of importance): technological factors (logical and physical), human factors, organizational factors.

In particular, the study identified ten main risk factors for distributed information systems, which can be displayed as follows (in order of importance and criticality of potential consequences): Factor_21, Factor_20, Factor_6, Factor_10, Factor_11, Factor_12, Factor_33, Factor_27, Factor_14, Factor_8. Nearly all respondents ranked factors related to lack of cybersecurity policy, lack of protection mechanisms against network attacks, violations of authentication and session management, violations of access control, and use of components with known vulnerabilities as the most important. These factors should be identified in the process of assessing and managing information security risks and monitored in the future.

Analysis of the most common categories of risk management mechanisms and measures used in modern distributed systems has shown that most protection controls are used frequently and are important mechanisms for preventing and minimizing potential risks. The generalization of the survey responses, according to the main groups of information security controls of modern distributed information systems in terms of frequency and effectiveness of their use, showed that most of the respondents identified controls responsible for the proper and effective use of cryptography and public key infrastructure, logical and physical access control, operational security and compliance with information security policies as important and most common in practice.

The results of the study can be used by managers and information security engineers to assess the importance and probability of potential risks and further prevent and minimize their consequences, as well as build tools for identifying and analyzing the risks of distributed systems based on qualitative, quantitative and intelligent methods.

6. References

[1] Andrew S. Tanenbaum, Maarten Van Steen Distributed Systems: Principles and Paradigms, Prentice Hall of India; 2nd edition (January 1, 2007)

[2] Henry K. Risk management and analysis / Kevin Henry // Information Security Management Handbook / Edited by Harold F. Tipton, Micki Krauze. - 6th edition. - Boca Raton: Auerbach Publications, 2017. - Part 1, Section 1.4, Ch. 28. - P. 321-329.

[3] Medvedeva, E. Organizatsyia integrirovannogo riskmenedzhmenta v organizatsyi // Vestnik nauki i obrazovaniya. 2020. № 24-4 (78). P. 23-26.

[4] Kanatov, M. Expert systems for information security management and audit. Implementation phase issues / M. Kanatov, L. Atymtayeva, B. Yagaliyeva // 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS). – 2014. doi: 10.1109/scis-isis.2014.7044702

[5] Trofymova N. Sovremennye tendentsyi korporativnogo risk-menedzhmenta v sisteme obespecheniya ekonomicheskoi ustoichivosti promyshlennykh predpriyatiy // UPRAVLENIE T. 8 № 2 / 2020. Mezhotraslevoi menedzhment. P. 30-38.

[6] State of Enterprise Risk Management 2020 Survey // ISACA, CMMI Institute. - 2019. - https://www.isaca.org/-/media/info/state-of-enterprise-risk-management-survey/index.html

[7] Rot A. IT Risk Assessment: Quantitative and Qualitative Approach // Proceedings of the World Congress on Engineering and Computer Science, 2008. - p. 1073-1078.

[8] 2019 Global Cyber Risk Perception Survey // Marsh, Microsoft. - 2019. - https://www.microsoft.com/security/blog/wp-content/uploads/2019/ 09/Marsh-Microsoft-2019-Global-Cyber-Risk-Perception-Survey.pdf.

[9] Konev I. Informatsyonnaya bezopasnost predpriyatiya. / I. Konev, A. Beliaev - SPb.: BKhVPeterburg, 2003.

[10] Chang, L.-Y. Applying fuzzy expert system to information security risk Assessment - A case study on an attendance system [Text] / L.-Y. Chang, Z.-J. Lee // 2013 International Conference on Fuzzy Theory and Its Applications (iFUZZY). - 2013. doi: 10.1109/ifuzzy.2013.6825462

[11] Xin Y. et al. Machine learning and deep learning methods for cybersecurity //IEEE access. – 2018. – Vol. 6. – P. 35365-35381.

[12] NIST Special Publication 800-30 Rev A. Risk Management Guide for Information Technology Systems, Gary Stoneburner, Alice Goguen, and Alexis Feringa, July 2002.

[13] Palko D., Myrutenko L., Babenko T., Bigdan A.: Model of Information Security Critical Incident Risk Assessment. 2020 IEEE International Conference on Problems of Infocommunications Science and Technology, PIC S and T 2020, 2021, pp. 157–161, 9468107.

[14] ISO/IEC 27001:2013. Information technology - Security techniques - Information security management systems - Requirements. 2013.

[15] ISO/IEC 27002:2013. Information technology - Security techniques - Code of practice for information security controls. 2013

[16] ISO/IEC 27005:2011. Information technology - Security techniques - Information security risk management. 2011.

7. Appendix

Appendix 1. Tables data

Table A1
Top Security Risks Factors of Modern Distributed Information Systems Based on Researchers

1 Category		2 №	3 Risk factors
Technological factors	Logical (software)	Factor_1	Insecure applications use
		Factor_2	Inadequate patch management
		Factor_3	API vulnerabilities and breaches
		Factor_4	Technical flaws and errors during system design
		Factor_5	Insufficient logging and monitoring
		Factor_6	Broken authentication and session management
		Factor_7	Unapproved third-party software use
		Factor_8	Use of unlicensed software solutions with undeclared capabilities
		Factor_9	0-day vulnerabilities and errors associated with the development of information technology
		Factor_10	Broken access control[3]
	Physical (hardware)	Factor_11	Using outdated hardware and components with known vulnerabilities
		Factor_12	Servers and network appliances security misconfiguration
		Factor_13	Low reliability of the set of hardware and software components, lack of a recovery plan, and periodic backups
		Factor_14	Weak endpoints and network perimeter protection
		Factor_15	Unmanaged IoT and mobile devices
		Factor_16	The imperfection of the organizational structure of the information security, the need for frequent reconfiguration of the information security or its individual parts
		Factor_17	The possibility of information leakage and sensitive data exposure using technical channels
		Factor_18	Insufficient physical access control
		Factor_19	Unauthorized use of the organization's assets
		Factor_20	Lack of protection mechanisms against external network attacks
Organizational factors		Factor_21	Lack of a cybersecurity policy
		Factor_22	Non-compliance with the requirements of standards at the stage of design of the system
		Factor_23	Non-compliance with information security requirements during system exploitation
		Factor_24	Lack of control over information security incidents
		Factor_25	Lack of top management commitment support and involvement
		Factor_26	Lack of security audits

[3] Lack of differentiation of user rights and controlled area access

Table A1 (continued)

1	2	3
	Factor_27	Lack of antivirus protection policy
	Factor_28	Weak potential to apply existing protection technologies
	Factor_29	Inconsistencies between the infrastructure and the adopted security measures
	Factor_30	The inability to provide the proper level of support and comprehensive development of security systems
Human factors	Factor_31	Actions of unreliable employees
	Factor_32	Unintentional mistakes of service personnel
	Factor_33	Privilege abuse
	Factor_34	The essential list of persons with access to protected information
	Factor_35	Lack of personnel awareness (especially about phishing/social engineering)
	Factor_36	Lack of information security training
	Factor_37	A severe shortage of cybersecurity professionals
	Factor_38	Insufficient passwords hygiene
	Factor_39	Personnel access to potentially dangerous objects in the external network
	Factor_40	Data loss or theft controls lack

Table A2

Key Security Controls of Modern Distributed Information Systems

1 №	2 Security control	3 Description
Control_1	Information security policies	controls responsible for implementing and verifying compliance with information security policies
Control_2	Organization of information security	controls responsible for the organizational component of information security measures and the distribution of responsibilities; creation of a management system for initiating and monitoring the implementation and operation of information security in the organization
Control_3	Personnel and human resources security	controls designed to regulate the work of personnel and contractors, identifying their responsibilities for information security both at the stage of the working process and upon dismissal
Control_4	Asset management	controls related to the inventory of company assets, classification of processed information, and media management
Control_5	Logical access control	controls responsible for restricting access to information and information processing facilities, access control policy, rights management for authorized users to systems and applications

Table A2 (continued)

1	2	3
Control_6	Cryptography	controls responsible for the proper and effective use of cryptography and public key infrastructure (PKI) to protect the confidentiality, reliability, and integrity of information
Control_7	Physical and environmental security	controls related to the management and prevention of unauthorized physical access, loss, damage, theft or compromise of assets and interruption of the organization's activities, as well as the definition of safe zones, entry controls, equipment security, “clear desk” and “clear screen” policies
Control_8	Operational security	a set of controls for ensuring the correct and secure work of processing information means that combines such activity as change management, backup, monitoring, logging and activity logs management, tracking the installed software and detecting malicious software, monitoring and eliminating identified vulnerabilities
Control_9	Communications security	controls related to network security, network services, information transmission, and messaging
Control_10	System acquisition, development, and maintenance	controls that define security requirements and protection mechanisms in development and support processes
Control_11	Supplier relationships	controls regarding relationships with third parties and contractors, protecting the organization's valuable assets that are available to them and ensuring an agreed level of information security and service delivery under agreements with suppliers
Control_12	Information security incident management	controls related to incident management, events, and information security vulnerabilities, reporting on identified violations, defining responsibilities, response procedures, and collecting evidence
Control_13	Information security aspects of business continuity management	controls that are necessary to ensure business continuity planning, verification and ongoing audit procedures, the availability of resources and information processing facilities, the use of resiliency and reliability principles to ensure security
Control_14	Compliance	controls that require compliance with legal and contractual requirements to avoid breaches of statutory, regulatory, or contractual obligations related to information security, procedures for protecting intellectual property, personal data, and assessing information security at all stages of the life cycle

Table A3

Mean Score for Each Risk Factor in Lifecycle of Modern Distributed Information Systems

Category		№	N	Mean	Std. Deviation	% percent
Technological factors	Logical (software)	Factor_1	23	3.043478	0.824525	60.8695
		Factor_2	23	2.826087	0.886883	56.5217
		Factor_3	23	3.695652	0.764840	73.9130
		Factor_4	23	3.739130	0.540824	74.7826
		Factor_5	23	2.782609	0.795243	55.6521
		Factor_6	23	4.217391	0.735868	84.3478
		Factor_7	23	2.695652	0.764840	53.9130
		Factor_8	23	3.869565	0.694416	77.3913
		Factor_9	23	3.826087	0.777652	76.5217
		Factor_10	23	4.173913	0.650327	83.4782
		Total	**23**	**3.4869564**	**0.7435418**	**69.7391**
	Physical (hardware)	Factor_11	23	4.130435	0.625543	82.6087
		Factor_12	23	4.086957	0.668312	81.7391
		Factor_13	23	3.434783	0.895752	68.6956
		Factor_14	23	3.913043	0.792754	78.2608
		Factor_15	23	2.869565	0.868873	57.3913
		Factor_16	23	1.913043	0.733178	38.2608
		Factor_17	23	3.434783	0.843482	68.6956
		Factor_18	23	3.086957	0.733178	61.7391
		Factor_19	23	2.869565	0.757049	57.3913
		Factor_20	23	4.304348	0.634950	86.0869
		Total	**23**	**3.4043479**	**0.7553071**	**68.0869**
Organizational factors		Factor_21	23	4.391304	0.656376	87.8260
		Factor_22	23	2.434783	0.787752	48.6956
		Factor_23	23	2.608696	0.782718	52.1739
		Factor_24	23	2.521739	0.845822	50.4347
		Factor_25	23	2.739130	0.810016	54.7826
		Factor_26	23	2.391304	0.838783	47.8260
		Factor_27	23	3.956522	0.824525	79.1304
		Factor_28	23	2.913043	0.596432	58.2608
		Factor_29	23	2.782609	0.795243	55.6521
		Factor_30	23	3.043478	0.638055	60.8695
		Total	**23**	**2.9782608**	**0.7575722**	**59.5652**
Human factors		Factor_31	23	2.782609	0.599736	55.6521
		Factor_32	23	2.304348	0.764840	46.0869
		Factor_33	23	4.000000	0.738549	80
		Factor_34	23	2.869565	0.548083	57.3913
		Factor_35	23	3.391304	0.782718	67.8260
		Factor_36	23	3.000000	0.603023	60
		Factor_37	23	3.347826	0.884652	66.9565
		Factor_38	23	3.260870	0.688700	65.2174
		Factor_39	23	2.086957	0.668312	41.7391
		Factor_40	23	3.782609	0.599736	75.6521
		Total	**23**	**3.0826088**	**0.6878349**	**61.6521**

Table A4

Testing the Hypothesis about the Relationship between Variables Using Spearman's Correlation Coefficient

		Fact_21	Fact_20	Fact_6	Fact_10	Fact_11	Fact_12	Fact_33	Fact_27	Fact_14	Fact_8
Fact_21	Correlation_Coefficient	1.000	.065	-0.95	-.056	-.055	.205	.000	.118	.340	-.251
	Sig. (2-tailed)		.770	.665	.801	.803	.344	1.00	.592	.112	.248
	N	23	23	23	23	23	23	23	23	23	23
Fact_20	Correlation_Coefficient	.065	1.000	.119	.244	.134	.110	.107	-.164	.066	-.528**
	Sig. (2-tailed)	.770		.588	.262	.544	.616	.628	.453	.765	.010
	N	23	23	23	23	23	23	23	23	23	23
Fact_6	Correlation_Coefficient	-.095	.119	1.000	.005	.262	.071	-.069	.123	-.019	-.132
	Sig. (2-tailed)	.665	.588		.983	.227	.749	.755	.577	.932	.548
	N	23	23	23	23	23	23	23	23	23	23
Fact_10	Correlation_Coefficient	.056	.244	.005	1.00	.423*	-.234	.388	-.094	-.309	-.157
	Sig. (2-tailed)	.801	.262	.983		.044	.283	.067	.668	.152	.476
	N	23	23	23	23	23	23	23	23	23	23
Fact_11	Correlation_Coefficient	-.055	.134	.262	.423*	1.000	-.145	.193	.089	.040	-.038
	Sig. (2-tailed)	.803	.544	.227	.044		.510	.377			
	N	23	23	23	23	23	23	23	23	23	23
Fact_12	Correlation_Coefficient	.207	.110	.071	-.234	-.145*	1.000	.086	.183	.354	-.361
	Sig. (2-tailed)	.344	.616	.749	.283	.510		.695	.404	.098	.091
	N	23	23	23	23	23	23	23	23	23	23
Fact_33	Correlation_Coefficient	.000	.107	-.069	.388	.193	.086	1.000	-.200	-.401	.270
	Sig. (2-tailed)	1.000	.628	.755	.067	.377	.695		.360	.058	.214
	N	23	23	23	23	23	23	23	23	23	23
Fact_27	Correlation_Coefficient	.118	-.164	.123	-.094	.089	.183	-.200	1.000	.177	-.026
	Sig. (2-tailed)	.592	.453	.577	.668	.685	.404	.360		.418	.906
	N	23	23	23	23	23	23	23	23	23	23
Fact_14	Correlation_Coefficient	.340	.066	-.019	-.309	.040	.354	-.401	.177	1.000	-.415*
	Sig. (2-tailed)	.112	.765	.932	.152	.856	.098	.058	-.418		.049
	N	23	23	23	23	23	23	23	23	23	23
Fact_8	Correlation_Coefficient	-.251	-.528**	-.132	-.157	-.038	-.361	.270	-.026	-.415*	1.000
	Sig. (2-tailed)	.248	.010	.548	.476	.863	.091	.214	.906	.049	
	N	23	23	23	23	23	23	23	23	23	23

Appendix 2. Frequency histograms for key risk factors

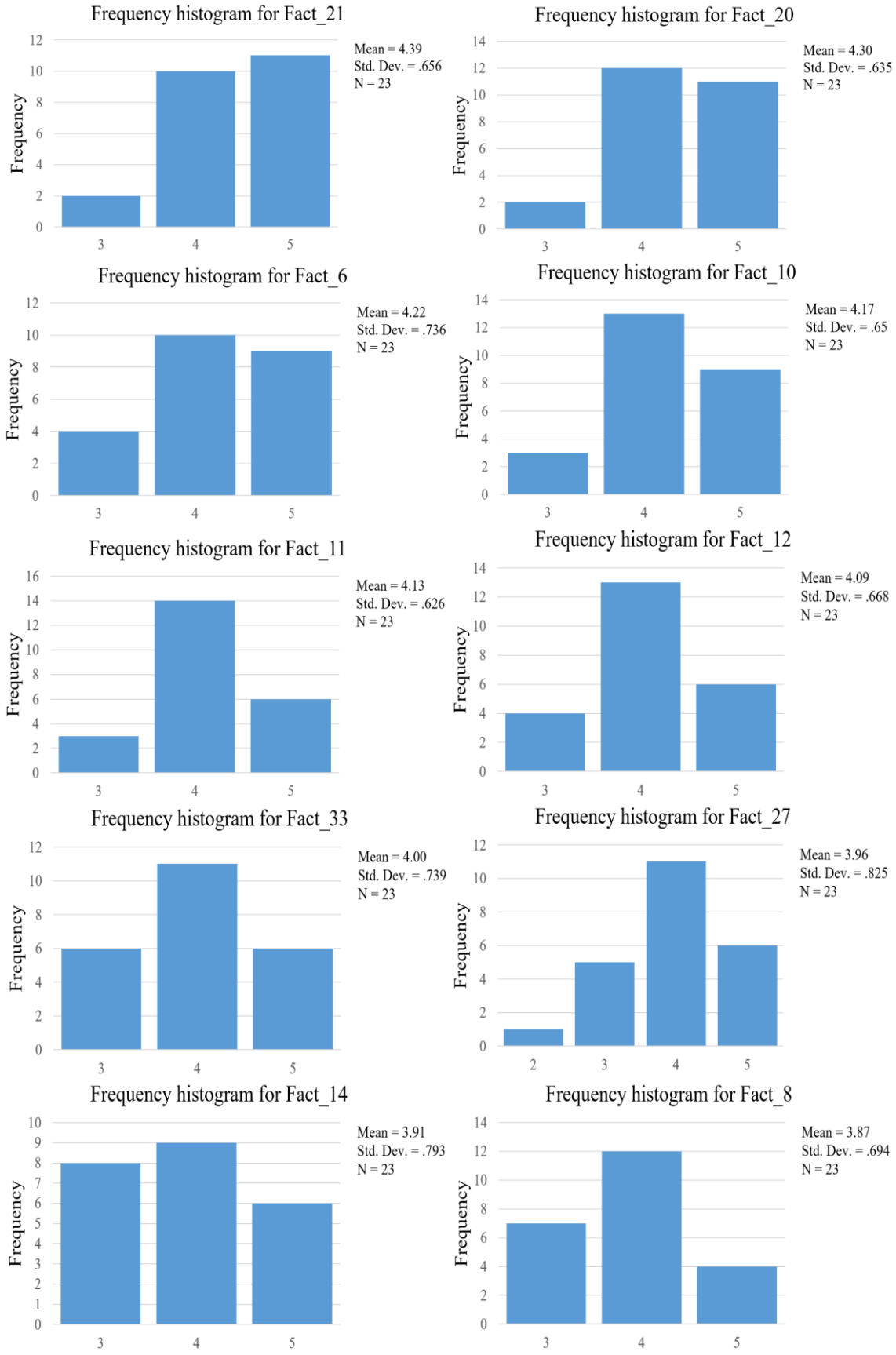

Figure A1: Frequency histograms for key risk factors