UDC 004.056

A Method for Assessing Risk with Accounting for the Structure of Threat and Vulnerability Relationships in a Complex System

Viktoriia Polutsyhanova[1], Serhii Smyrnov[1]

[1] *Educational and Research Institute of Physics and Technology*
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Abstract
A generalized method for assessing risks using structural analysis of relationships between threats and vulnerabilities in the system is described. The use of Q-analysis is proposed as a basic method for describing the structure of the system, which reveals complex relationships between vulnerabilities and threats, and allows for refining risk assessments. An improved risk assessment formula based on Bayesian assessment is developed using the assumption of compatibility of threat implementation depending on the profile of attacks on the system.

Keywords: cyber system, risk assessment, vulnerabilities, threats, Bayesian methods, Q-analysis, simplex complex, cybersecurity.

Introduction

Recently, the popularity of risk assessment methods and approaches has increased significantly. Regulatory acts and standards [6] are increasingly being adopted that contain risk assessment rules and also require the use of these methods in various areas. This contributes to the development and improvement of relevant methodological approaches. Taking into account the above, this trend is expected to continue to be significant [5].

The main purpose of any risk analysis is to provide guidance for decision-making, especially when it comes to cybersecurity. When a decision making related to risk, a risk assessment process that includes an understanding of the sources of risk is useful. The use of a risk assessment method can solve a variety of problems, including global problems, such as the location of production facilities, as well as technical problems regarding the specifics of the system's functioning, including human and organizational problems.

Risk assessment should provide more objective data, which ultimately will help to find a compromise between increasing profits and minimizing negative consequences. This is an iterative search that leads to continuous improvement in the decision-making process and, in the ideal case, contributes to increased efficiency in cybersecurity.

Risk assessment is also used in the quality assessment system. The implementation of a quality system should facilitate the use of different methods and sources of information, that is, information of a certain degree that satisfies user requests. Similar to risk, the quality level of an institution can be derived from the institutional environment and the goals of the institution. In this context, the institutional environment has a significant impact on the organization's tolerance for risk to achieve goals.

The process of risk assessment and management can be divided into several stages: defining the structure, identifying risks, analyzing the likelihood and impact of risks, assessing risks, and finally responding to risks. This study mostly concerns the risk assessment stage but at the same time complements the methodology of the entire risk management cycle in systems.

1. Classification of risk models in complex systems

This paper examines the complex dependencies between vulnerabilities and threats.

This relationship is cascading or non-binary in nature, compared to graphs. Previous research [1] has highlighted the structural analysis and classification using Q-analysis of these relationships. Methods for transforming from graphs to simplex complexes are also provided to better represent the relationship between vulnerabilities and threats. Research shows that traditional approaches and methods for calculating risks are not enough to describe the risks of systems with a complex structure. Therefore, improvements are needed for a more complete and qualitative representation of the structural characteristics of information and cyber systems.

In [3-5], existing methods for calculating risk and finding source data for them are analyzed. In the case of incidents, when assessing risk, difficulties arise in assessing the probability and scale of damage from their implementation.

The research results suggest a method for calculating risks that can take into account the complex relationships between system elements and vulnerabilities that arise during their life cycle of cyber systems.

Let us consider a few examples. We use the classical Bayesian formula (1) to calculate risk:

$$\boldsymbol{R = \sum_i \ p_i V_i} \qquad (1)$$

where p_i is the probability of an event (vulnerability realization), V_i is the amount of loss in the event of an event, $i = \overline{1,n}$ is the vulnerability index.

This formula is often very abstract and does not take into account the specifics of the collected input data. As mentioned earlier, the main drawback of this approach is that the total amount of losses due to incidents caused by the use of vulnerabilities and threats that occur frequently and inconsistently does not lead to significant losses and, therefore, may be equal, in terms of risk level, to events with low probability but significant losses.

In order to take into account, the specifics of the relationships between vulnerabilities, it is recommended to use a model of interaction and influence between vulnerabilities based on Q-analysis, taking into account the structural features that arise when building the modeled complex. The general scheme for forming a Bayesian risk assessment based on losses and probabilities is shown in Figure 1.

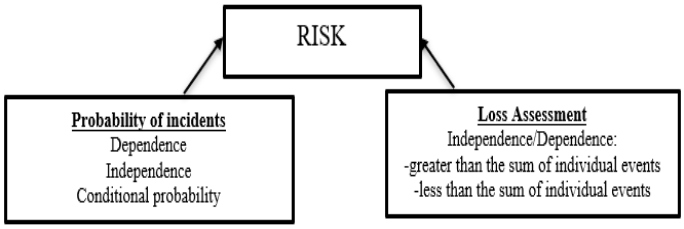

Figure 1: Scheme of risk assessment refinements

During the study, the estimates of the corresponding probabilities and losses were refined. Classical formulas for average losses take into account the sum and ratio of adverse events, but often the assessment of the probability of threat realization is simplified for the case of individual incompatible events.

In practice, compatible realizations of vulnerabilities/threats may arise, which complicates the assessment of probability due to the appearance of compatible probabilities and conditional probabilities. Also, the risk assessment is refined by revising the assessment of the number of losses in the event of the occurrence of compatible events.

It is obvious that if individual vulnerabilities are independent, then the total loss from their joint realization is equal to the sum of losses from individual incidents. At the same time, considering the case of realization in which compatible vulnerabilities arise, the level of loss is considered as a function that depends on the losses caused by each of these events. As a result, losses may increase or decrease compared to the linear assessment. For example, if the events do not have a significant impact on the system, a more effective way to overcome the negative consequences is to simultaneously process several events.

However, in some cases, the total damage caused by a cyber incident may be greater than the sum of the losses caused by individual cyber incidents. For example, if the realized risk causes a system failure. This is a situation when a single event causes corresponding damage to the system, but the system is viable, but due to the simultaneous occurrence of a large number of such events, the cyber system may lose its ability to function.

Below is a classification of potential cases for constructing such an assessment for risk and loss. In the case when the vulnerabilities of the events are incompatible, classical approaches to risk

assessment are used. The numbering of formulas in this case corresponds to the level of complexity of the connection [7].

1.0. Formula (2) for calculating risk in the case when the vulnerabilities of the events are incompatible:

$$R = \sum_i p_i V_i, \qquad (2)$$

where $\sum_i p_i < 1$ та $\exists p_0 > 0$ – is the probability of the system functioning without losses, V_i is the losses from events.

This formula is the same as in classical risk assessment. The only limitation is that the sum of probabilities is less than 1, since there may be terms in which the sum of losses is zero. This means that even if there is a probability of some losses, their total sum is not significant compared to other possible total losses.

The following formulas represent a combination of different types of dependencies between events.

2.0. Formula (3) for calculating risk in the case of pairwise compatibility of vulnerabilities in the implementation of events and their independence (probabilistic and by losses):

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_i p_j (V_i + V_j) \qquad (3)$$

This formula describes the case when, in addition to independent events (losses), there are also pairwise compatible realizations of events. That is, there is a probability that some or all events will occur in pairs. At the same time, since the events do not depend on each other, the joint probability is calculated as a product, and the loss function is calculated as a sum.

2.1. Formula (4) for calculating risk in the case of pairwise compatibility of vulnerabilities in the realization of events, their probabilistic dependence and their independence in terms of losses:

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_{\{i,j\}} (V_i + V_j) \qquad (4)$$

This formula also reflects pairwise compatibility. That is, the possibility of some events occurring simultaneously, so in the second term the loss function reflects how total losses are formed (the sum of individual losses), but at the same time their probabilistic dependence is taken into account. That is, individual events are related to each other and are the result of other events. For example, when one event occurs as a result of another event or when several of these events occur simultaneously. In this case, the probability is not equal to the product of individual components, but is calculated using conditional probabilities or determined by experts.

2.2. Formula (5) for calculating risk in the case of pairwise compatibility and independence of vulnerabilities in the event of the occurrence of events by probability and their dependence on losses:

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_i p_j V_{\{i,j\}}, \qquad (5)$$

$$V_{\{i,j\}} \neq V_i + V_j$$

2.3. Formula (6) for calculating risk in the case of pairwise compatibility of vulnerabilities in the realization of events, their probability dependence and their dependence on losses:

$$R = \sum_i p_i V_i + \sum_{\substack{i,j,i\neq j \\ V_{\{i,j\}} \neq V_i + V_j}} p_{\{i,j\}} V_{\{i,j\}}, \qquad (6)$$

The next case is the case of not only the probability dependence of events, but also the dependence on losses. That is, when two events occur simultaneously, the total loss is not equal to the sum of the losses in individual events.

The loss function is used if there is a simple case of risk realization, in which the total loss is less than the sum of the losses in individual events. At the same time, if the corresponding risk is critical for the cyber system, the total loss may exceed the amount of loss of an individual event. In addition, taking into account the dependence of events on their probabilities, risk assessment becomes more laborious, but takes

into account the specific properties and structure of the cyber system.

Below are formulas describing more advanced risk assessment methods that take into account the triple and larger compatible dependence of risk on probability and loss. The formulations of these formulas are similar to the formulas of double-even compatibility given above, so their description will be omitted.

Of greater interest are formulas that reflect the calculation of risk for fully compatible events and their dependence on losses, taking into account the probability of their occurrence (in this case, n is the multidimensional complexity of the relationships between threats and vulnerabilities and can achieve full connectivity between them).

n.0. Formula (7) for calculating risk in the case of full compatibility of vulnerabilities in the implementation of events and their complete independence from probability and losses:

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_i p_j (V_i + V_j) + \cdots + \prod_{i=1}^{n} p_i \sum_{i=1}^{n} V_i \quad (7)$$

n.1. Formula (8) for calculating risk in the case of full compatibility, independence of vulnerabilities in the implementation of events by probability and their dependence on losses:

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_i p_j V_{\{i,j\}} + \cdots + \prod_{i=1}^{n} p_i V_{\{i,j,\dots,z\}} \quad (8)$$

n.2. Formula (9) for calculating risk in the case of complete compatibility of vulnerabilities in the realization of events, their independence in terms of losses and dependence on probability:

$$R = \sum_i p_i V_i + \sum_i p_{\{i,j\}} (V_i + V_j) + \cdots + p_{\{i,j,\dots,z\}} \sum_{i=1}^{n} V_i \quad (9)$$

n.3. The most general formula (10) for calculating risk when all possible combinations of vulnerabilities in the realization of events are available:

$$R = \sum_i p_i V_i + \sum_{i,j,i\neq j} p_{\{i,j\}} V_{\{i,j\}} + \sum_{i,j,k,i\neq j\neq k} p_{\{i,j,k\}} V_{\{i,j,k\}} + \cdots + p_{\{i,\dots,n\}} V_{\{i,\dots,n\}} \quad (10)$$

For each leaf of the structural threat tree, a version of the risk calculation formula is used, so that for each simplex the sum of probabilities will be less than one. But with the complexity of each level of adjacency, the number of terms in the risk formula increases as the tree progresses. To clearly structure risk systems, a structural tree based on Q-analysis is needed [2].

The last formula for risk calculation describes the most complex options when events with high probability and interdependence are calculated. That is, any event can cause a cascade of various combinations of other events. At the same time, since these events are compatible, this means that such implementations are allowed.

The specified formula covers situations when independent loss realization events are considered, as well as options when the total loss is higher or lower than the amount of loss for a separate event. It covers various options for possible implementation of measures. Calculation according to this formula is quite laborious. At the same time, in practice, the probability of some events may be insignificant or even zero, or the losses themselves are not significant or zero, so they can be ignored. Based on this, in the multiplicative formula, the corresponding terms will be zero and can be neglected.

The proposed formula is usually applied to a simplex in a complex. Terms with more complex structures make a smaller contribution, because their probabilities are very small, and their consequences are smaller. Taking into account the above, a low probability of an event significantly reduces the impact on the possible development of the system, although these events can lead to significant losses for the

organization. That is, the compatibility of events is achieved due to the existence of simplex connections between them.

2. Using structural analysis for risk assessments

Structural features arising from simplex complexes allow us to take into account the compatibility between vulnerabilities when assessing risk. In this context, compatibility means that the same vulnerabilities can arise (be introduced) simultaneously or separately, depending on the circumstances. Based on this, such situations should be taken into account when calculating the risk, since the more complex the connections between individual vulnerabilities, the greater the contribution of individual components to the overall risk.

But the peculiarity of this situation is that when calculating the overall risk of the ICS, it is necessary to subtract the risk of connections between simplex chains from the overall risk. Since they are taken into account both when calculating the risk of individual subsystems and when calculating the total risk (an example of total probability). That is, the impact of this connection is taken into account several times. That is, the impact of such connections is taken into account several times. When the complex is considered as a whole, there is duplication of connections between vulnerabilities that belong to different simplices, but through which the simplices are "glued together". Therefore, it is necessary to eliminate these duplications when calculating the overall systemic risk.

The calculation of the total risk begins with formula (11), which takes into account all the leaves of the structural tree, which are simplices of different dimensions:

$$R = \sum_{i} r_i + \cdots + \sum_{\{i,\dots,k\}} r_{\{i,\dots,k\}}, \qquad (11)$$

where $i, \dots, k = \overline{1, N}$, i N - number of simplexes.

Each leaf of the structural tree (which is a separate simplex chain) corresponds to a separate part of the general risk calculation formula. But as the structure tree is traversed, the complexity of the chain at each connection level and the corresponding number of members in the risk formula increase. Therefore, for an unambiguous calculation of the full risk assessment, it is necessary to apply a simple complex composition procedure proposed in the second section of the study.

It is proven that the structural features of the proposed simplex complex allow taking into account the compatibility between threats and vulnerabilities when assessing risk. In this context, compatibility means that individual vulnerabilities can be initiated (introduced) simultaneously or independently, depending on different circumstances. At the same time, it was found that the more complex the relationships between individual vulnerabilities, the greater the impact of compatible components on the overall risk.

When calculating the overall risk of the system, it is also necessary to take into account the “glues” (connections) between simplex chains. These adjacencies are simplexes by definition, so the risk associated with this chain is formed from parts of the two chains. The sum of the risks of the two chains is calculated, and the risk of the simplex connection is subtracted from it, since it is duplicated from each individual chain.

Considering “k” chains, simplex connections occur simultaneously for all chains together. It is clear that in order to balance the large number of connections, we need to subtract the risk from the simplex of the adjacency, multiplied by (k-1), from the total risk in the chain.

Based on this, the total risk (through the chain) must be adjusted for each vertex of the structure tree to account for such duplications.

Information about the adjacency structure is included in the local map and the structure graph of the complex and can be used to calculate the risk directly.

In general, the risk calculation is similar to the well-known “tree folding” algorithm of decision theory, but with “inclusions and exclusions” and the corresponding multiplicities of adjacency.

Therefore, the above-described general risk assessment for the complex system as a whole is calculated by formula (12):

$$R_{total.} = R - R^* \qquad (12)$$

In the above formula, R is the risk in the system, defined by the simplex (the component

of the original complex that "glues" the complex together). However, the risk is exaggerated because it includes repetitions that arise when considering several combinations of chains of simplexes connected to each other by tangent vertices, edges, faces, etc. To compensate for the redundancy, we calculate *R**.

Using a generalized example, we will consider a risk calculation that takes into account the structure of the system. Let us assume that the system has a system structure as in Figure 2.

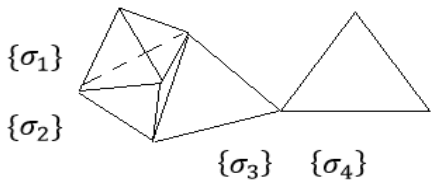

Figure 2: Simplex complex for the system structure

In this case $\{\sigma_1\}$, $\{\sigma_2\}$, $\{\sigma_3\}$, $\{\sigma_4\}$ – simplices, which make up the complex structure of the system. Each of them consists of elements that generate it, as well as connections of different arity.

$\{\sigma_1\}$ – consists of four elements, let's call them e_1, e_2, e_3, e_4, which are interconnected by third−level connections, that is, each of them is simultaneously connected to the others and simultaneously forms a spatial connection in the form of a tetrahedron.

$\{\sigma_2\}$ has a similar structure, consists of 3 e_5, e_2, e_3, e_4. These simplices are connected to each other through a second-order simplex, i.e. a triangle with vertices e_2, e_3, e_4.

$\{\sigma_3\}$ represents a simplex of dimension 2, consisting of elements e_4, e_5, e_6. Through a one-dimensional relationship consisting of e_4, e_5, $\{\sigma_3\}$, he is connected with $\{\sigma_2\}$.

The last element of the complex is $\{\sigma_4\}$, which consists of e_7, e_8, e_6. It is connected through a 0-dimensional connection e_6 with $\{\sigma_3\}$.

In order to correctly calculate the risk level, data is required for the calculation $R_{\sigma_1}, R_{\sigma_2}, R_{\sigma_3}, R_{\sigma_4}$, and also with R_{e_2,e_3,e_4}, R_{e_4,e_5}, R_{e_6}.

The final values are calculated as risks of the corresponding simplexes, so they have certain features.

Let us assume that for our example we have all the necessary data. In this case, the general form of the risk assessment formula (13) will be as follows:

$$R = \sum_{i=1}^{4} R_{\sigma_i} - R_{e_2,e_3,e_4} - R_{e_4,e_5} - R_{e_6}. \quad (13)$$

This type of risk assessment formula is more practical because it takes into account the details of the system structure and allows for a more accurate level of assessment. Note that if two or more simplices are connected simultaneously due to the presence of q-connected simplices defined by vertices, they must be multiplied by a factor of m-1, where m is the number of simplices connected by this q-connection.

After calculating the risk of each subsystem, element and performing an overall risk assessment of the entire system, it is possible to draw conclusions about how individual vulnerabilities affect the overall risk of the entire system. The proposed approach allows for more effective prioritization and ranking of security issues.

The advantage of this approach is that it is more balanced and structurally sound than a conventional risk matrix and, accordingly, Q-analysis is used to calculate a given risk, better reflecting the structure of the system. Using the inverse algorithm allows us to build simplex complexes. The direct algorithm allows us to determine the priorities of vulnerabilities.

With proper analysis, it is possible to find the probability distribution of vulnerabilities in the system. The latter is useful when the statistical distribution of vulnerabilities and threats is not determined. It should be noted that the indicators that characterize potential losses are quite subjective due to the complex mechanisms for obtaining reliable information.

This is explained by the fact that this is partly confidential information, which is often based on expert assessments. The latter requires additional research to ensure an appropriate level of security for the system using event scenarios that implement vulnerabilities through unauthorized intrusion, such as hacker attacks on information systems.

3. Generalized method for calculating risk assessment

Below we summarize all the approaches developed during the study and present a complete method for calculating risk assessment of systems with a complex structure. In this study, it is assumed that each vulnerability system has a complex structure, which is the result of existing and potential threats to the system. We present all the stages of the resulting generalized method.

Stage I.

Collection of data on system vulnerabilities and threat structure.

Based on the available information, associations and dependencies between threats and vulnerabilities of the system are determined and the corresponding incidence matrix is created.

Data can be presented in the form of structural graphs (simplified complexes), regular graphs or descriptions of interaction models between elements.

Stage II.

Synthesis of simplicial complexes.

The method of constructing a system complex using the incidence matrix [1] is applied.

Stage III.

Determination of structural features of the simplicial complex using Q-analysis methods.

The structure tree, local map and hierarchy of descendants are constructed. These properties will be used at the next stage of the methodology.

Stage IV.

Classification of threats/vulnerabilities in the complex based on Q-analysis.

Based on the identified structural features of the system such as q-connection, q-connectivity and hierarchy of descendants, the classification of threats/vulnerabilities in the symplectic complex is carried out. This classification can be used instead of an ordinal scale, for example, as a level of threat/vulnerability criticality in the absence of reliable estimates.

Stage V.

Probability distribution and determination of the size of losses.

Based on the profile of attacks on the system, a probability distribution for threats is formed. Using expert methods, estimates of losses from vulnerabilities and threats are determined depending on their combination.

Stage VI.

Calculation of risk estimates for the system.

Collapse of the structural tree. Based on local maps and the structural graph, a formula for calculating the overall risk of the system is synthesized.

Using local maps and the structural tree, we build a formula for calculating the overall risk of the system:

- for each leaf of the tree at any level of connectivity, the risks for the corresponding simplex are calculated (each risk is partial, but its calculation is not trivial);
- when moving along the structural tree, it leads to the fact that individual simplexes are connected into chains, that is, simplexes with different degrees of q-connection are “glued”.

In the general system risk formula, the correction for adjacency glues is subtracted as the value of the calculated risk for this glue. That is, if two 3-dimensional simplexes are glued together by a 2-dimensional simplex, you must calculate the risk for each simplex separately, add the risks of the glued simplexes and subtract the value of the glue risk. This is necessary in order not to take into account the risk from the glue again.

By examining the entire structural tree, using the general loss function and the risk assessment formula, it is possible to obtain the form of a formula for calculating the risk assessment of systems with a complex structure. This approach to calculating additive indicators is more objective if the form of the structural tree and local maps are additionally taken into account.

Thus, this method has a wide range of applications and can be applied not only to the assessment of systems and vulnerabilities, but also to other structurally complex systems if they use the complex systems synthesis procedure, which takes into account the "inclusion-exclusion" of impacts from different components.

Conclusions

This study develops a generalized risk assessment method for systems with a complex structure. Depending on the compatibility of the implementation of the threat structure components, different methods for calculating the average risk function in the system are considered. A formula for calculating the loss function is given, taking into account the probabilistic properties of compatible implementations of vulnerabilities in the system.

It is proposed to calculate the risk of the entire complex using an additive formula that corrects for redundancy due to the large number of connections between simplexes. The proposed formula has a polynomial form due to the compatibility of threat and attack profiles.

A method for constructing a Bayesian risk assessment formula is proposed that takes into account the structure of simplex complexes created on the basis of a system of connections between vulnerabilities and threats.

At the same time, the method allows for fairly simple analytical studies to identify the maximum and minimum risk, as well as conditions under which a high level of correction for "gluing" may occur.

References

[1] Polutsyganova V. I., Smirnov S. A. Methodology for constructing the main metrics of q-analysis and their application. System Research and Information Technologies. 2019. No. 3. P. 76 – 88. URL: https://doi.org/10.20535/srit.2308-8893.2019.3.07 (date of access: 10.11.2023).

[2] Atkin R. H. Mathematical structure in human affairs [Текст] / Atkin. – London: Heinemann Educational Books, 1973. – 143 c.

[3] Wald A. Statistical Decision Functions. The Annals of Mathematical Statistics. 1949. Vol. 20, no. 2. P. 165–205. URL: https://doi.org/10.1214/aoms/1177730030 (date of access: 06.12.2023).

[4] Ferguson T. S. Mathematical statistics: A decision theoreticapproach. New York : Academic Press, 1967. – 396 p.

[5] Cox L. A. Risk analysis of complex and uncertain systems. Boston, MA: Springer US, 2009. URL: https://doi.org/10.1007/978-0-387-89014-2 (date of access: 10.09.2023).

[6] ISO/IEC 27005:2018, Information Security Risk Management, International Organization for Standardization, Geneva, Switzerland, 2018.

Polutsyganova, V. I. Risk assessment method based on analysis of the structure of connections between threats and vulnerabilities in cybersystems: dissertation Doctor of Philosophy: 125 – Cybersecurity and information protection / Polutsyganova Viktoriia Ihorivna. – Kyiv, 2024. – 207 p.