

Analysis of Features and Prospects of Application of Dynamic Iterative Assessment of Information Security Risks

Denis Berestov^a, Oleg Kurchenko^a, Yuri Shechblanin^b, Nataliia Korshun^c,
and Tetiana Opryshko^c

^a Taras Shevchenko National University of Kyiv, 24 Bohdan Hawrylyshyn str., Kyiv, 04116, Ukraine

^b State Enterprise "Ukrainian Special Systems," 83b Yuri Illenko str., Kyiv, 04119, Ukraine

^c Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska str., Kyiv, 04053, Ukraine

Abstract

The article is devoted to the approach to information security risk analysis. The factors influencing the risk analysis process are defined. In such a task there is always a prior probabilistic information about the implementation of threats, which may be changed after the receipt of new expert assessments or as a result of observation of relevant events. One way of "revision" of the relative acceptability of probabilistic models is Bayesian approach, the essence of which implies that the degrees of trust in possible probabilistic models to obtain data are considered. After the information has been received, the probabilities are re-evaluated. In the analysis of information security risks, probabilistic models of the studied systems are used. Probabilistic space of events in the field of information security is determined and in probabilistic space the probabilistic measure is set by this or that method. To solve this problem an artificial neural network can be used. As an alternative to Bayesian approach, the method of maximum function of likelihood can be considered, which is used in the statistical estimation of distribution parameters. Bayesian approach to solving problems has advantages, as many properties of estimates obtained using the likelihood ratio are not performed in the case of a small sample size. Applying Bayesian approach also helps to solve the question of mathematical methods of assessment of prior values that can take the parameters of information security risk. In the presence of a large amount of statistics, the wrong choice of a prior distribution of probabilities will not significantly affect a posterior one. In the absence of such data it is expedient to choose a distribution that minimally affects a posterior distribution. The estimation of probability of realization of threats to information security exploiting relevant vulnerabilities is obtained by using Bayesian network.

Keywords

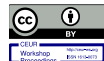
Risk; Bayesian approach, vulnerability, information system model, prognostication, neural network.

1. Introduction

Formulation of the problem. In conditions of increasing complexity of automated systems, issues of information security are becoming more and more important for the state and business. Particular attention is beginning to be paid to the analysis of information security risks as a necessary component of an integrated approach to information security. As a result, a large number of standards and approaches, the basic concepts and definitions in this field are characterized by plurality.

The most appropriate definition of risk for most practical applications of information security is given in the ISO 27005 Standard and the BS7799 Standard. According to ISO 27005, "Information security risk is a potential opportunity to exploit asset vulnerabilities or a group of assets of a specific

Cybersecurity Providing in Information and Telecommunication Systems, January 28, 2021, Kyiv, Ukraine
EMAIL: berestov@ukr.net (A.1); kuro1@ukr.net (A.2); shechblanin@ukr.net (B.3); n.korshun@kubg.edu.ua (C.4); t.opryshko@kubg.edu.ua (C.5)
ORCID: 0000-0002-3918-2978 (A.1); 0000-0002-3507-2392 (A.2); 0000-0002-3231-6750 (B.3); 0000-0002-4055-1494 (C.4); 0000-0002-9282-0182 (C.5)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

threat to the detriment of the organization.” As follows from this definition, risk is a complex quantity defined as a function (or functional) of a number of other quantities. Difficulties in conducting risk analysis are directly related to difficulties and errors in the analysis of risk components.

In addition to organizational problems we can identify the following main issues:

- Obvious lack of information about the components of risk and their ambiguous properties.
- Creating a model of information system.
- The duration of the process and the rapid loss of relevance of the evaluation results.
- Aggregation of data from various sources including statistics and expert assessments.
- The need to involve specialists in risk analysis.

In this regard, the methods of continuous audit and analysis of information security risks, which are actively being developed, have acquired special relevance. Together with modern models of management of information systems, information security management systems, monitoring and security analysis, these methodologies allow to quickly and efficiently build and develop the organization’s information security system. The main types of audit of security in the preparation of terms of reference for the design and development of information protection system and - after its implementation - assessing the level of its effectiveness and the main stages of the audit are discussed in detail in the work [9].

The system of continuous dynamic audit and risk analysis allows specialists to conduct iterative risk assessment taking into account the available data on the business landscape, relevant information on technologies that are used or intended for implementation. A special role in the continuous risk analysis should be played by the function of forecasting information security risks. By automating the process of accounting for threats associated with the emergence of new vulnerabilities in the standard software, formalizing changes in the business landscape and information system, aggregation of data from different sources, an environment can be created that allows professionals to make reports on the level of security of a particular information system, based on a series of consecutive reports compiled in a short period of time.

Processing the data using the methods of statistical forecasting will determine the optimal set of countermeasures taking into account “future risks” and thus increasing the efficiency of their implementation. Therefore, it is necessary to synthesize an approach to obtaining a quantitative estimate and information security risk management in an automated system, taking into account:

- Possibility of aggregation of heterogeneous data.
- The opportunity to learn in the process of work and refine the assessments obtained at previous stages.
- The ability to work with deliberately inaccurate data.
- The ability to automate most decision-making processes.

To create such an environment requires the construction of a model of an automated system, which in itself is a complex task that usually implies significant simplifications. In order to solve these problems it is necessary to synthesize an automated system that allows to fully or partially automate the process of describing the operating environment and derivation of risk values.

Analysis of recent research and publications. The issue of applying Bayesian approach to information security risk analysis is affected in a number of works by Ukrainian and foreign authors, among which we can highlight [4–6]. For example, in [4] an algorithm in decision - making in operational management of information security tools based on Bayesian networks of trust and the method of analysis of hierarchies are proposed, which allows to predict the state of protected resource and prevent the development of dangerous situations in a timely manner.

The use of Bayesian networks for information security risk assessment is substantiated in [5]. For calculation of conditional probabilities the approach on the basis of trees of attack as Bayesian special networks is proposed . Bayesian networks provide an efficient way to combine historical quantitative data with qualitative ones. It is possible to use conditional probabilities to take into account the interdependence of vulnerabilities. Also the model of using Bayesian networks to assess the damage from information security events is suggested.

In the work [6] an attempt was made to compare Bayesian approach to risk analysis of information security with the approach based on the calculation of fuzzy logic (the integral of Choquet).

In the works [7] and [8] it is shown that an artificial neural network can be used for solution of this problem. The most common approach is to train neural network so that it implements the nonlinear function of discrimination, which provides a direct division of the observed input vectors into classes. More general and a promising approach is to teach the neural network so that the output values of the system were the posterior probabilities of belonging of input data to the set classes.

It was also shown in [7] that it is possible to build a neural network, which after its training will directly obtain estimates of the conditional probability $p(x|y)$. It is also possible to define methods of training of such networks and methods of evaluating the results obtained.

The purpose of the article. The aim of the article is to highlight the application of Bayesian approach and the apparatus of Bayesian networks for the analysis of information security risks in the implementation of security breaches by exploitation of the greatest vulnerability.

2. Theoretical Fundamentals of Research

Bayesian approach in risk analysis. If it is necessary to estimate quantitative value of the risk, it is accepted to use the definition given in the BS 7799 Standard. Accordingly, definition of risk R can be given by the following expression:

$$R = P \cdot C \quad (1)$$

where P is the probability of realization of the threat and C is the magnitude of the consequences.

In its turn the probability of realization of the threat is the product of the probability of realization of vulnerability to the probability of exploitation of this vulnerability. In this interpretation of the probability of a random event, the realization of the threat can be considered as a random variable. Given the complexity of calculating the probability of threats based on analytical approach there is a need to assess the probability P for risk analysis tasks expertly or based on statistics on the frequency of implementation of this class of threats for a given type of automated systems at a given time interval. In this case, the implementation of the second approach requires the use of a fairly large statistical material, accumulated over a certain period of time.

The methodologies considered during the analysis do not contain direct instructions as to the methods of analysis of the statistical data used. In this regard it is expedient to study ways of solving this problem in terms of mathematical statistics. The tasks similar to information security risk analysis arise in different areas. Among them are risk analysis in the field of finance and management of software projects.

The main mathematical methods used in these fields are traditional school of statistical inference described in the works of Newman, Pearson, Fisher and others. A number of works offer alternative methods of analysis. Methods based on Bayesian approach are of particular interest.

As mentioned above, the task of risk analysis always has a prior probabilistic information on the implementation of threats that may be changed after receiving new expert assessments or as the result of observing relevant events related to assets that confirm or refute a prior information. Many statistical tasks regardless of the methods of their solution have a common property: before a specific data set is obtained, as potentially acceptable for the situation under study, several probabilistic models should be considered. Once the data is obtained, there appears some form of knowledge of the relative acceptability of these models.

One way to "revise" the relative acceptability of probabilistic models is Bayesian approach, which is based on Bayes' theorem. Bayesian approach in the science of management is formulated in [1] as a scientific discipline, which is based on the principle of maximum use of available prior information, its regular review and reevaluation taking into account the obtained sample data on the phenomenon or process being investigated. Such a review is interpreted as learning, and the process of management itself in the Bayesian approach is understood as a process of learning (adaptation).

The essence of Bayesian approach is that it considers the degrees of trust in the possible probabilistic models to obtain data. Degrees of trust are presented in the form of probabilities. After obtaining the information using Bayes' theorem, probabilities are revalued, i.e. new values of probabilities are calculated, reflecting degrees of confidence in probabilistic models taking into account the newly

obtained data. Probabilistic models are used in the analysis of information security risks of the studied systems.

Their essence is that the probabilistic space of events is determined in the field of information security and in probabilistic space a probabilistic measure is set by one or other method. In the conditions of probabilistic models the observed parameter of information security system can be considered as a random variable Y_c with the density of probability $P(y)$. Based on these observations, a conclusion is made about the probability distribution of random variable. Then according to Bayesian formula:

$$P(\theta|y) = \frac{P(y|\theta) \cdot P(\theta)}{P(y)} \quad (2)$$

or

$$P(y|\theta) = \frac{P(\theta y) \cdot P(y)}{P(\theta)} \quad (3)$$

Bayesian approach has found application in many areas, including cryptographic analysis, in the field of risk analysis in various modern systems of artificial intelligence designed to work in conditions of uncertainty. The most obvious is the use of this approach when using the apparatus of Bayesian networks. The probabilistic model, called Bayesian network, is built using many variables and their probability dependences, which are presented in the form of directional acyclic graph. The vertices of the graph are variables or hypotheses, and the edges represent conditional dependence of variables or hypotheses. There are effective methods [2, 3] of creation (calculation) and Bayesian network training.

3. Research Results

Let's consider the application of Bayesian approach to information security risk analysis in more detail. In [10] it is shown that the model of security violator, which is constantly adjusted on the basis of new knowledge gained about the capabilities of the violator, and changes in the protection system based on the analysis of the causes of violations that have occurred, will influence these causes and more precisely determine the requirements for the information security system.

We use this model of information security violator. In the case of implementation of network security threat an attacker develops an attack script that uses one vulnerability. In this case, the security breach occurs through the exploitation of the largest vulnerability. If several vulnerabilities are equivalent, one of them is chosen arbitrarily. Let's consider the following Bayesian network. Its interpretation in terms of risk analysis can be next.

Let us denote the possible threat to the information security of the system as T , the probability of successful implementation of the threat—as $p(T)$, the existing vulnerabilities—as U_1, \dots, U_n , and the probability of successful exploitation of vulnerabilities— $p(U_1), \dots, p(U_n)$ respectively. The conditional “attack potential” value is denoted by A . The variable $A[0,1]$ is a random variable with distribution $f(A)$. Values $u_1, \dots, u_n \in [0,1]$. We will say that the threat U is not realized if $A > p(U_i)$, and is realized if $A < p(U_i)$. For ease of writing, we denote as U_i the fact of exploitation of the vulnerability, and \bar{U}_i is the absence of the fact of exploitation of the vulnerability. As part of the analysis of risks it is important to consider the likelihood of threats being realized through the exploitation of the relevant vulnerabilities.

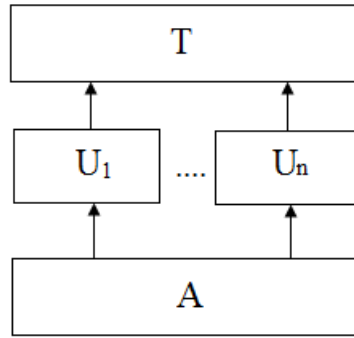


Figure 1: Bayesian network for the implementation of information security threats

Probability of events in the common space in the conditions of the model of independent threats is defined by the expression:

$$P(T, (U_1, \dots, U_n), A) = P(T|(U_1, \dots, U_n))P(A) \prod_{i=1}^n P(U_i|A) \quad (4)$$

Performing marginalization to P (T) we obtain:

$$P(T) = \int P(A|(U_1, \dots, U_n))x \times [\int \prod_{i=1}^n P(U_i|A)f(A)dA]d(U_1, \dots, U_n) \quad (5)$$

From the dichotomous nature of variables and conditions (6)

$$A \leq p(U_i) \Rightarrow p(U_i) = 0 \text{ and } A > p(U_i) \Rightarrow p(U_i) = 1 \quad (6)$$

it follows that

$$P(U_i|A) = 0 \text{ if } A \leq p(U_i) \text{ and } P(U_i|A) = 1 \text{ if } A > p(U_i) \quad (7)$$

$$P(U_i|A) = 0 \text{ if } A \leq p(U_i) \text{ and } P(U_i|A) = 1 \text{ if } A > p(U_i)$$

This conclusion can be represented graphically for some values of p (Uj) and p (Uk) as follows. According to this approach we obtain

$$\int \prod_{i=1}^n P(U_i|A)f(A)dA = \min(1 - U_1, \dots, 1 - U_n) \quad (8)$$

$$\int P(\neg U_j|A) \prod_{i \neq j} P(U_i|A)f(A)dA = \max\left(0, U_j - \sum_{i \neq j} U_i\right) \quad (9)$$

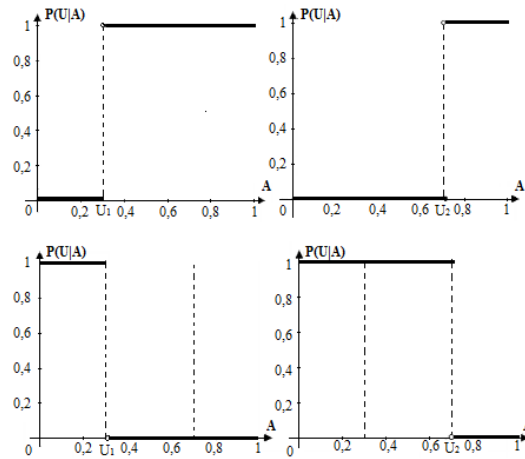


Figure 2: The values of $p(U_j)$ and $p(U_k)$ under condition A

Substituting these expressions in (6) we obtain an estimate of the probability of implementation of threats to information security. This conclusion can be applied to practical tasks of information security risk assessment. Let's consider the solution of this expression in the case of uniform distribution of $f(A)$ for two threats.

$$\int P(U_1|A)P(U_2|A)P(A)dA = \min(1 - u_1, 1 - u_2) \quad (10)$$

$$\begin{aligned} \int P(U_1|A)P(\neg U_2|A)P(A)dA &= \max(0, u_2 - u_1) = \\ &= \max(0, (1 - u_1) + (1 - (1 - u_2)) - 1) \quad (11) \end{aligned}$$

$$\begin{aligned} \int P(\neg U_1|A)P(U_2|A)P(A)dA &= \max(0, u_2 - u_1) = \\ &= \max(0, (1 - (1 - u_1)) + (1 - u_2) - 1) \quad (12) \end{aligned}$$

Substituting the obtained probabilities into expression (6), we obtain:

$$\begin{aligned} P(T) &= P(T|U_1, U_2) \min(1 - u_1, 1 - u_2) + \\ &+ P(T|U_1, \neg U_2) \max(0, u_2 - u_1) + P(T|\neg U_1, U_2) \max(0, u_1 - u_2) \quad (13) \end{aligned}$$

After simple transformations, expression 14 takes the following form:

$$\begin{aligned} P(T) &= P(T|U_1, \neg U_2)(1 - u_1) + P(T|\neg U_1, U_2)(1 - u_2) + \\ &+ (P(T|U_1, U_2) - P(T|\neg U_1, U_2) - P(T|U_1, \neg U_2)) \times \\ &\times \min(1 - u_1, 1 - u_2). \quad (14) \end{aligned}$$

As an alternative to Bayesian approach, the method of maximum function of likelihood, which is used in the statistical estimation of distribution parameters, can be considered. Bayesian approach to solving the assigned problems has advantages, as many properties of estimates obtained using the likelihood ratio, are not performed in the case of a small sample size.

4. Conclusions and Prospects of Further Research

Applying Bayesian approach also helps to address mathematical issues of methods for estimating the prior values that can take risk parameters of information security. An important feature is that in the presence of a large volume of statistics, the wrong choice of prior probability distribution will not essentially affect the posterior probability. However, what is especially true for solving problems of analysis of information security risks, in the absence of such data it is advisable to choose distribution that minimally affects the posterior distribution (so-called non-informative distribution).

These conclusions are very important for the tasks of analysis of information security risks and can be very useful for solving the problem of their iterative dynamic analysis.

The use of artificial neural networks to solve the assigned problem allows us to provide a number of important properties of the system as well, including provision of the ability to teach the system in the process of functioning and its adaptation to different conditions of functioning. When using them, there is also no need for the previous detailed modeling of the automated system. The direction of further research is aimed at the use of approximate estimates of a posterior probability of realization of events.

5. References

[1] L. I. Lopatnikov, *Economic and Mathematical Dictionary: Dictionary of Modern Economic Science*, 5-th ed., Moscow, Russia: Delo, 2013.

[2] F. V. Jensen, T. D. Nielsen, *Bayesian Networks and Decision Graphs*. New York, NY: Springer New York, 2016.

[3] S. Russell, J. Pearl, *Bayesian Networks*, University of California, Tech. Rep. R-277 November 2000, 2014.

[4] D. N. Pogorelov, *Protection of information resources of the enterprise on the basis of multiagent technology*, Ph.D. thesis, Ufa, 2006.

[5] R. M. Alguliev, Y. N. Imamverdiev, S. A. Derakshande, *Information security risk assessment using Bayesian networks*, *Telecommunications* 6 (2017) 30-34.

[6] E. Lavrentyev, M. V. Timonin, *Comparative analysis of approaches to information security risk modeling, based on the theory of fuzzy sets and Bayesian networks*, *Information Technology Security* 3 (2014) 97-101.

[7] C. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag. New York, 2016.

[8] G. Dreyfus, *Neural Networks*, Berlin/Heidelberg: Springer-Verlag, 2015.

[9] Y. V. Roy, N. P. Mazur, P. M. Skladannyi, *Audit of information security is the basis of effective protection of the enterprise*, *Cybersecurity: Education, Science, Technique* 1 (2018) 86-93, doi: 10.28925/2663-4023.2018.1.8693.

[10] Y. Shcheblanin and D. Rabchun, *Mathematical model of information security's threat agent*, *Cybersecurity: Education, Science, Technique* 1 (2018) 63-72, doi: 10.28925/2663-4023.2018.1.6372.