TEKA Kom. Mot. i Energ. Roln. – OL PAN, 2011, 11A, 129-136

MODELLING OF DISCRETE RECOGNITION AND INFORMATION VULNERABILITY SEARCH PROCEDURES

Valerie Lahno, Alexander Petrov

Department of Computer Systems and Networks,
Volodymyr Dahl East-Ukrainian National University, Lugansk, Ukraine
AGH University of Science and Technology, Poland

Summary. The article to contain results of the researches, allowing to raise level of protection of the automated and intellectual information systems enterprises (AIS). The article discusses the use of discrete procedures to detect threats for information resources.

Key words: information security, threat detection, discrete process.

INTRODUCTION

Information security management has become a critical and challenging business function because of reasons such as rising cost of security breaches, increasing scale, scope and sophistication of information security attacks, complexity of information technology (IT) environments, shortage of qualified security professionals, diverse security solutions from vendors, and compliance and regulatory obligations.

The sophistication and effectiveness of cyber attacks have steadily advanced. These attacks often take advantage of flaws in software code, use exploits that can circumvent signature-based tools that commonly identify and prevent known threats, and social engineering techniques designed to trick the unsuspecting user into divulging sensitive information or propagating attacks. These attacks are becoming increasingly automated with the use of botnets - compromised computers that can be remotely controlled by attackers to automatically launch attacks. Bots (short for robots) have become a key automation tool to speed the infection of vulnerable systems [Ahmad D. 2005, Chi S.-D. 2001, Gorodetski V. 2002, Knight J. 2002, Templeton S. 2000, Xiang Y. 2004].

RESEARCH OBJECT

Mission-critical information systems (MCIS) are understood as the electronic communication development objects, by means of which collection, processing, storage and transmission of information are performed with the purpose to ensure the handling processes. Their exceedance of allowable values may lead to the malfunction or their endamagement.

To evaluate security of such a system, a security analyst needs to take into account the effects of interactions of local vulnerabilities and find global vulnerabilities introduced by interactions. This requires an appropriate modeling of the system. Important information such as the connectivity of elements in the system and security related attributes of each element need to be modeled so that analysis can be performed. Analysis of security vulnerabilities, the most likely attack path, probability of attack at various elements in the system, an overall security metric etc. is useful in improving the overall security and robustness of the system. Various aspects which need to be considered while deciding on an appropriate model for representation and analysis are: ease of modeling, scalability of computation, and utility of the performed analysis. The analysis of the protection of information systems and automated control systems for transport companies has yielded the following results (period 2008 -2010), fig. 1, 2.

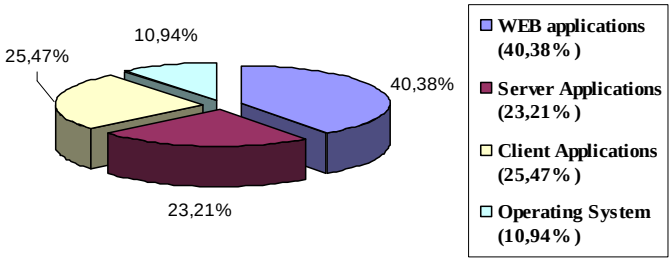

Fig. 1. Statistics application vulnerabilities AIS

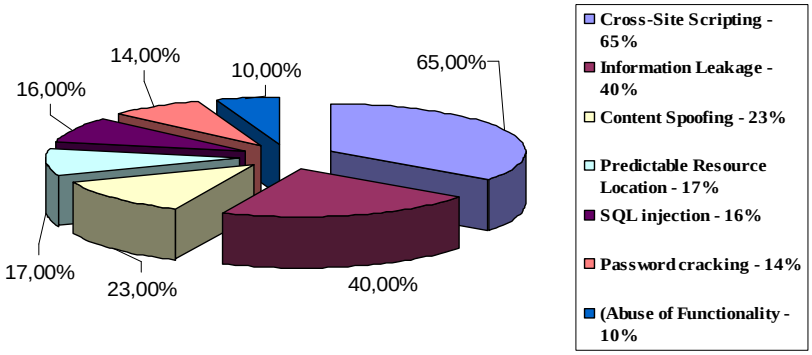

Fig. 2. The probability of detecting vulnerabilities of different types

The decision of questions of complex maintenance of security and stability of functioning of the automated systems (AS) in the conditions of unauthorized access (UNA), including, influences of computer attacks, demands the system analysis and synthesis of possible variants of construction of means of counteraction UNA means. At complex formation it is necessary to co-ordinate and inter connect functions and parameters of the EXPERT, protection frames of the information from UNA, anti-virus means, gateway screens, the communication equipment, the general and special software and perspective means of counteraction to computer attacks [Chapman C. 2003].

The main peculiarity of the concerned recognition and software and network vulnerability search procedures, which are later called discrete or logical procedures, is the possibility of obtaining a result without any information about functions of character meaning distribution and on availability of little training samples. The knowledge of metrics in the space of objects' description is not needed also. In this case a binary function of value proximity should be determined for each of the characters, which allows distinguishing the objects and their sub descriptions [Baskakova L. 1981, Vayntsvayg M. 1973].

The main task of discrete recognition and vulnerability search procedures (DRVSP) building is search of informative sub descriptions (or description fragments) of objects.

We consider informative objects to be the objects that reflect certain regularities in description of objects used for training, that is presence or, vice versa, absence of these fragments in the object, which is being considered, allows attributing it to one of classes. The fragments that are met in descriptions of one class objects and cannot be met in descriptions of other classes' objects are considered to informative in DRVSP. The regarded fragments as a rule have a substantial description in terms of designing information safety systems (ISS).

RESULTS OF RESEARCH

The main task of discrete recognition and vulnerability search procedures (DRVSP) building is search of informative sub descriptions (or description fragments) of objects.

We consider informative objects to be the objects that reflect certain regularities in description of objects used for training, that is presence or, vice versa, absence of these fragments in the object, which is being considered, allows attributing it to one of classes. The fragments that are met in descriptions of one class objects and cannot be met in descriptions of other classes' objects are considered to informative in DRVSP. The regarded fragments as a rule have a substantial description in terms of designing information safety systems (ISS).

A notion of an elementary classifier is introduced by building discrete recognition and vulnerability search procedures for information safety systems. An elementary classifier is understood as a fragment in a description of a training sample. A certain multitude of elementary classifiers with preset properties are built for each

$(KL_1,...,KL_l)=(B_{p_{a1}},...,B_{p_{al}})$ class. As a rule, the classifiers, which are used, can be met in descriptions of one class objects and cannot be met in descriptions of other classes' objects, thus describing only some training objects of the class. On the other hand, sets of character values not used in descriptions of any training objects of the class characterize all objects of this class and are more informative form this perspective. That is why so actual is the question of constructing discrete recognition and vulnerability search procedures based on the principle of "nonreoccurance" of character legitimate values' sets, fig. 3, 4.

Another problem is presence of objects which are on borderline between classes $(KL_1,...,KL_l)=(B_{p_{a1}},...,B_{p_{al}})$ among the study samples of objects. Each of such objects is not "typical" for its class, as it resembles to descriptions of objects belonging to other classes. Presence of untypical objects extends the length of fragments used to distinguish objects belonging to different classes. Long fragments are less frequent in new object, thus extending the number of unrecognized objects.

The necessity of building effective realizations for discrete recognition and vulnerability search procedures is directly connected to problems of metric (quantitative) characters of informative fragments' multitudes. The most important and technically complex are the problems of obtaining asymptotical estimates for typical number values of (impasse) covering and the length of integer matrix (impasse) covering and also the problems of obtaining analogical estimates for permissible and maximum conjunctions of a logical function, which are used for synthesis of circuit hardware-based ISS solutions.

There is, as a rule, no reliable information about the structure of *PA* multitude available while solving tasks connected with projecting an effective AIS information safety system, that's why having built a discrete recognition and vulnerability search procedures algorithm we cannot guarantee its high performance on new objects different from $\{sp_{a1},...,sp_{am}\}$. Nevertheless, if the training samples are quite typical for the considered multitude of objects, than the algorithm that makes infrequent mistakes in studies will show acceptable results with unknown (not included in training samples) objects also. In this connection correctness of discerning algorithm is the problem that should be paid great attention. The algorithm is considered to be correct if it discerns all the training samples correctly.

The simplest example of a correct algorithm is the following: the considered object sp_{an} is compared to descriptions of every training sample $\{sp_{a1},...,sp_{am}\}$. In case if the sp_{an} object's description coincides with a description of a sp_{an} training sample, the sp_{an} object is attributed to the same class as the $\mathrm{sp}_{\mathrm{ai}}$ object. In other case the algorithm declines to recognize the object. There is no difficulty noticing that though the foregoing algorithm is correct, it is not able to discern any object which description does not coincide with description of any training sample.

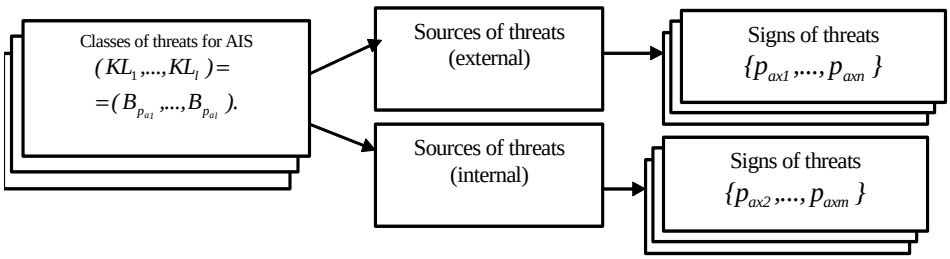

Fig. 3. The structure of the classification of “Sources of Threats”

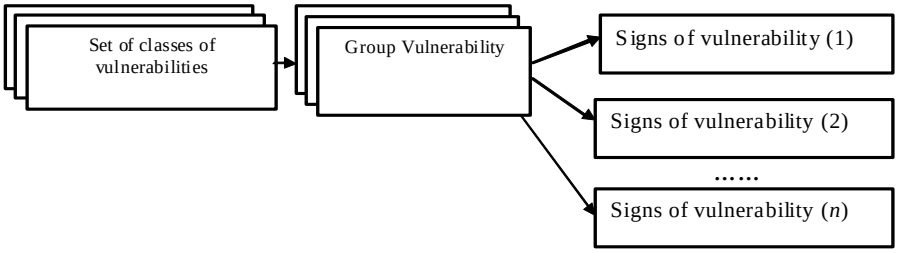

Fig. 4. The structure of the classification of “vulnerability”

Let’s introduce the following symbols. Let NP_{p_a} stand for a set of r_{p_a}, $r_{p_a} \le MI$ different integer-valued characters of $\{p_{aj_1},..., p_{aj_r}\}$ kind. Proximity of $sp'_a=(\alpha p'_{a1},\alpha p'_{a2},...\alpha p'_{aMI})$ and $sp''_a=(\alpha p''_{a1},\alpha p''_{a2},...\alpha p''_{aMI})$ belonging to *PA* by the NP_{p_a} set of characters we will estimate by the following value

$$BN(sp'_a, sp''_a, NP_{pa}) = \begin{cases} 1, & if \quad \alpha p'_{j_{ti}} = \alpha p''_{j_{ti}} \ the \ value \ of \ ti = 1,2,..., r_{p_a}, \\ 0 & otherwise \ . \end{cases} \quad (1)$$

Thus, the schematic circuit of estimation algorithm building for information safety systems is the following. The whole range of different $NP_{p_a}=\{p_{aj_1},..., p_{a_{MI}}\}$, $r_{p_a} \le MI$ type sub multitudes is picked out inside the $\{p_{a_1},...,p_{a_{jMI}}\}$ character system. Later the picked sub multitudes are named reference multitudes of the algorithm, and their whole range is designated by ΩMI.

Further let us set the following parameters:

- po_{sp_a} is a parameter characterizing significance of a sp_{ai}, *i= 1, 2,..., PA* target (object);

- $po_{NP_{pa}}$ is a parameter characterizing significance of an object belonging to a reference multitude $NP_{p_a} \in \Omega MI$.

The considered object sp_{an} is compared to every training sample sp_{ai} of every reference multitude. A $\Gamma(sp_a, KL)$ estimation of sp_a object belonging to *KL* class is calculated for each vulnerability class of AIS *KL*, $KL \in \{KL_1, ..., KL_l\}$ in the following way:

$$\Gamma(sp_a, KL) = \frac{1}{|LW_{KL}|} \sum_{sp_{ai} \in KL} \sum_{NP_{Pa} \in \Omega MI} po_{sp_a} \cdot po_{NP_{pa}} \cdot BN(sp_a, sp_{ai}, NP_{p_a}), \quad (2)$$

where: $|LW_{KL}| = |KL \cap \{sp_{a1}, ..., sp_{aMI.}\}|$.

The sp_{an} object is attributed to the class that has the highest estimate. In case if there are several classes with the highest estimate, discerning fails.

Let's regard the situation, when the objects of the considered PA multitude are described by the characters, each possessing values of the {0, 1,..., k_{p_a} - *1*} multitude. Let's associate the (σ_{DOP}, NP_{pa}) elementary classifier, where $\sigma_{DOP} = (\sigma_{DOP_1}, ..., \sigma_{DOP_r})$, NP_{pa} is a set of characters numbered $j_1, ... j_{r_{pa}}$, with an elementary conjunction $\Re = p_{axj_1}^{\sigma_{DOP_1}} ... p_{axj_{r_{pa}}}^{\sigma_{DOP_{r_{pa}}}}$.

Let's show that building a multitude of $(KL_l) = (B_{p_{al}})$ class elementary classifiers for the models previously considered in the article adds up to finding permissible and maximum conjunctions of the characteristic $(KL_l) = (B_{p_{al}})$ class function, which is a double-valued logical function possessing different values for training samples of KL_l и $\overline{KL_l}$.

After completion of all the previously mentioned stages one can start the work on forming the model of information threats for all the information resources of the enterprise on the basis of the derived classifiers. The initial data for simulation are classes of vulnerabilities, threats and attacks, and also multitudes of AS attack realization means and categories (classes) of malefactors.

The problem of using proper characteristic functions was not considered in corpore within the bounds of this research, as there are different mathematical approaches to descriptions of characteristic functions, which can be found for each class of information attack targets. For example, the following methods are used for solving problems connected with simulating the speed of malicious software spreading, that is measuring the percentage of infected computers within the network:

- models based on changed systems of differential equation, formulated in classic epidemiologic models;

- models based on calculation of Hamiltonian path length in the part of the analogous graph, where spreading is still possible;
- other.

REFERENCES

1. Ahmad D., Dubrovskiy A., Flinn X., 2005.: Defense from the hackers of corporate networks. Trudged. with angl. - 2th izd. M.: Companies AyTi; DMK - Press. 864 p.
2. Atighetchi M., Pal P., Webber F., Schantz R., Jones C., Loyall J., 2004.: Adaptive Cyberdefense for Survival and Intrusion Tolerance // Internet Computing. Vol. 8, No.6. p.25-33.
3. Atighetchi M., Pal P.P., Jones C.C., Rubel P., Schantz R.E., Loyall J.P., Zinky J.A., 2003.: Building Auto-Adaptive Distributed Applications: The QuO-APOD Experience // Proceedings of 3rd International Workshop Distributed Auto-adaptive and Reconfigurable Systems (DARES). Providence, Rhode Island, USA. p.74-84.
4. Baskakova L., Guravlev Y., 1981.: Model of recognizing algorithms with the representative sets and systems of supporting great numbers// Gurn. Vich. matem. and matem. Fiz. 21-5. p. 1264-1275.
5. Chapman C., Ward S., 2003.: Project Risk Management: processes, techniques and insights. Chichester, John Wiley. Vol. 1210.
6. Chi S., Park J., Jung K., Lee J., 2001.: Network Security Modeling and Cyber At-tack Simulation Methodology//LNCS. Vol. 2119.
7. Goldman R., 2002.: A Stochastic Model for Intrusions//LNCS. Vol. 2516.
8. Gorodetski V., Kotenko I., 2002.: Attacks against Computer Network: Formal Grammar-based Framework and Simulation Tool. RAID 2000//LNCS. Vol. 2516.
9. Harel D. Statecharts: A., 1987.: Visual Formalism for Complex Systems, Science of Computer Programming 8. p. 231-274.
10. Hariri S., Qu G., Dharmagadda T., Ramkishore M., Raghavendra C., 2003.: Impact Analysis of Faults and Attacks in Large-Scale Networks//IEEE Security & Privacy. p. 456-459.
11. Hatley D., Pirbhai I., 1988.: Strategies for Real-Time System Specification, Dorset House Publishing Co., Inc., NY. 930 p.
12. Keromytis A., Parekh J., Gross P., Kaiser G., Misra V., Nieh J., Rubensteiny D., Stolfo S., 2003.: A Holistic Approach to Service Survivability // Proceedings of ACM Workshop on Survivable and Self-Regenerative Systems. Fairfax, VA. p. 11-22.
13. Knight J., Heimbigner D., Wolf A.L., Carzaniga A., Hill J., Devanbu P., Gertz M., 2002.: The Willow Architecture: Comprehensive Survivability for Large-Scale Distributed Applications // Proceedings of International Conference Dependable Systems and Networks (DSN 02). Bethesda, MD, USA. p.17-26.
14. Lahno V., Petrov A., Skripkina A., 2010.: Construction of discrete recognition procedures, and vulnerability scan information. Information security № 2 (4). p. 5-13.
15. Lahno V., Petrov A., 2009.: Prevention from Penetration into Dynamic Database of Corporate Information Systems of Enterprises. Management of Organizatoon Finances, Production, Information. Bielsko-Biala. p. 282-290.
16. Smirniy M., Lahno V., Petrov A., 2009.: The research of the conflict request threads in the data protection systems. Proceedings of Lugansk branch of the International Academy of Informatization. № 2(20). V 2. 2009. p. 23-30.

17. Templeton S., Levitt K., 2000.: A Requires/Provides Model for Computer Attacks. Proc. of the New Security Paradigms Workshop. p. 274-280.
18. Vayntsvayg M., 1973.: Algorithm of teaching of pattern recognition is «Cora»// In kn.: Algorithms of teaching to pattern recognition. p. 82-91.
19. Xiang Y., Zhou W., Chowdhury M., 2004.: A Survey of Active and Passive Defence Mechanisms against DDoS Attacks. Technical Report, TR C04/02, School of Information Technology, Deakin University, Australia. p. 38-43.

МОДЕЛИРОВАНИЕ ДИСКРЕТНЫХ ПРОЦЕДУР РАСПОЗНАВАНИЯ УГРОЗ И ПОИСКА УЯЗВИМОСТЕЙ ИНФОРМАЦИИ

Валерий Лахно, Александр Петров

Аннотация. Статья содержит результаты исследований, позволяющие повысить уровень защиты автоматизированных и интеллектуальных информационных систем предприятия (АИС). В статье предложено использовать дискретные процедуры для выявления угроз информационным ресурсам.

Ключевые слова: информационная безопасность, обнаружения угроз, дискретный процесс.