Scientific and Practical Cyber Security Journal (SPCSJ) 2(3): 47-53 ISSN 2587-4667
Scientific Cyber Security Association (SCSA)

SIGNATURE AND STATISTICAL ANALYZERS IN THE CYBER ATTACK DETECTION SYSTEM

Toliupa S.[1], Druzhynin V.[2], Parkhomenko I.[3]
[1,3] Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[2] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
toliupa@i.ua, v_druzhinin@ukr.net, parkh08@ukr.net

ABSTRACT. The globalization of information exchange and the widespread introduction of information technologies in all spheres of society's life created the problem of protecting information processed in information systems from challenges and threats in the cybernetic space. The presence of important information in the functioning of the systems and objects of critical national infrastructures enables its usage by the negatively-minded elements and groupings for the implementation of unlawful actions in the cyber space by violating the integrity, availability and confidentiality of information, and inflicting damage on information resources and information systems. In this case, the possibility of using information technologies in the cybernetic space in the interests of carrying out military-political and power confrontation, terrorism and hacking cyber attacks is of a particular concern. The purpose of the article is to develop a system for recognizing cyber threats based on signature analysis, which would reduce the time of detection of an attack of a cyber defense system while the number and complexity of cyber attacks are increasing.

KEYWORDS: cyberspace, cyber attack, signature analyzer, decision-making system, cyber intrusion.

One of the main problems, which under the condition of globalization of information exchange and wide implementation of information technologies in all spheres of society's life support came up in all states of the world, is the problem of protecting information processed in information systems from challenges and threats in the cybernetic space. Possibilities of the cyberspace, rapid development and implementation of leading-edge information and telecommunication technologies provide the unprecedented opportunities for accumulation of data and its usage. The presence of important information in the functioning of the systems and objects of critical national infrastructures enables its usage by the negatively-minded elements and groupings for the implementation of unlawful actions in the cyber space by violating the integrity, availability and confidentiality of information, and inflicting damage on information resources and information systems. In this case, the possibility of using information technologies in the cybernetic space in the interests of carrying out military-political and power confrontation, terrorism and hacking cyber attacks is of a particular concern. The purpose of the article is to develop a system for recognizing cyber threats based on signature analysis, which would reduce the time of detection of an attack of a cyber defense system while the number and complexity of cyber attacks are increasing.

To compass this purpose, the following problems should be solved:

- to create a detection system of the aberrant behavior which is built upon the capability of the cyber attack detecting system to have a knowledge of some characteristics which describe the correct (or permissive) behavior of the object of observation;
- to develop a signature analyzer model which enables a cyber attack or cyber intrusion detection for critically important information structures;
- to develop a statistical analyzer on basis of the average-case analysis model and the root-mean-square deviation of network traffic settings.

As the world experience has showed, the most effective methodological approach for constructing of innovative intellectual cyber attack monitoring systems is the way to create a hierarchical multilevel structure of cyber attack detection at the beginning of their implementation. Furthermore, a hierarchical approach allows to solve difficult problems of the information protection process managing from cyberattacks in the distributed information systems (IS) as sequence of local tasks, coordinated with each other.

The estimation of threats of critically important systems involves two aspects: situational analysis and threats detection [1, 6, 7, 9]. The situational analysis is a detailed analysis of software settings functioning of the IS. While carrying out such an analysis it is necessary to organize similar data and estimate it separately according to each group. There is an example of such an analysis, presented in Fig. 1.

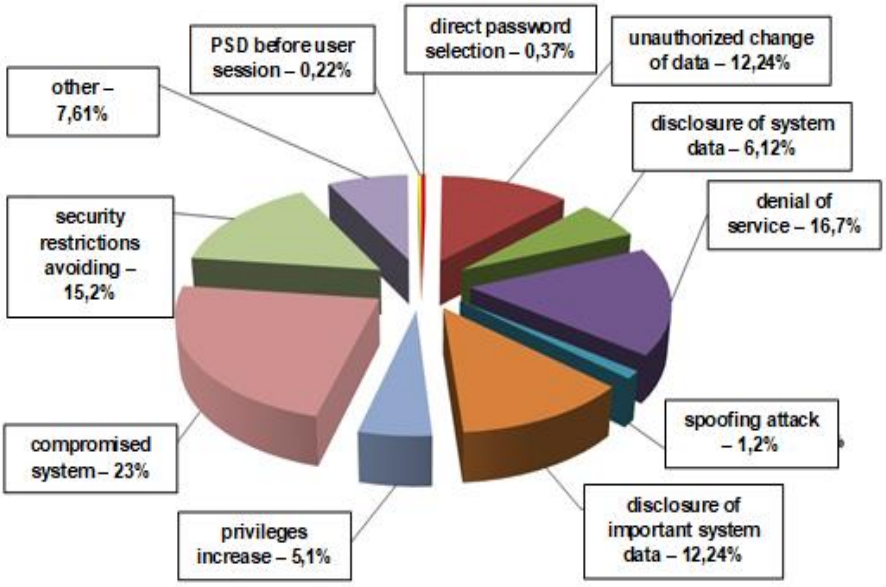

Fig.1

Nowadays, for the purpose of information systems protection it is necessary not only to develop private mechanisms protection, but also implement a system approach, which involves a complex of connected actions. The aim of any information safety related system is to prevent them from cyber attacks, protect the legal interests of a business entity from information security incidents, prevent from financial looting, dissemination, diseconomy, misrepresentation and destruction of information.

For today, systems of cyber intrusion and cyber attacks detection usually present program or machine-program solutions, which help automate actions control process taking place in the information system or in a network, and also analyze these actions directly to detect some cybersecurity warnings. As the number of different types and ways of organization of network hacking has increased for the recent years, CADS became a necessary component of a security infrastructure of most organizations.

In general, modern systems of intrusion and cyber attacks detection are far from ergonomic and effective solutions, according to the security. But the improvement of efficiency should be considered not only in the sphere of detection of improper activities on the infrastructure of secure information objects, but also according to everyday exploitation of these measures and to the saving of computing power and information resources of an owner of a security system.

If to talk straight about modules of data-processing, it should be remembered that every cyber attack signature in the system of information processing concerning a cyber attack is a basic element for detecting of most general actions – cyber attack phase detecting (the stage of its implementation). The definition of a *signature* itself is generalized to a final rule. On the contrary, each cyber attack is developed for the phase number of its development. The easier a cyber attack is, the easier it can be detected and there are more opportunities to analyze it.

The cyber attack scenario is a transition diagram which transits to an analogical diagram of the final determined automated device. Cyber attack phases can be described in the following way: ports testing; identification of program and machine tools; banner gathering; exploits usage; disorganization of network functioning with help of attacks for a customer service refusal; managing through backdoors; Trojans set searching; web proxies searching, presence signs removing and s.o (in appropriate cases – with different detalization level).

The benefits of such an approach are obvious - in the case of separate processing of various stages of cyber attacks, it is possible to recognize a cyber threat in the process of its preparation and formation, and not at the stage of its implementation, as in the existing systems. At the same time, the elemental basis for recognition can be a signature search, detection of anomalies, the use of expert methods and systems, trust relationships and other information methods in order to assess what is happening in the information environment. A general approach to analysis allows us to determine distributed (in all senses) cyber threats, both in logical and physical

space. The general scheme of event handling also allows searching for distributed cyber attacks by further data aggregating from different sources and constructing metadata about known incidents.

The cyber attack detection systems, like most modern software products, must meet a number of requirements. These are modern development technologies, orientation on the features of modern information networks and compatibility with other programs. To understand how to use CADS correctly, you need to clearly identify how they work and what their vulnerabilities are. If we do not take into account various non-essential innovations in the field of detection of cyber attacks, then we can safely assert that there are two main technologies of constructing the CADS.

The most widespread cyber threats to information resources can be considered as potentially possible cases of natural, technical or human-induced nature, which may lead to unwanted effects on the information system, as well as on the information stored therein. The emergence of a cyber threat, that is finding the source of actualization of certain events in the threat, is characterized by such an element as vulnerability. By integrating a variety of approaches, as well as suggestions for solving this issue, we believe that the following kinds of cyber threats to information security can be identified: disclosure of information resources; violation of their integrity; failure of the equipment itself.

Traditionally, CADS are classified according to two characteristics: the method of detection and the level of the system on which the protection is carried out. Despite the fact that these two classification features are most important in the selection of systems for detecting cyber attacks, there are still other characteristics that play an equally important role in the design of the CADS. After all, the safest solution can not be achieved by considering one or two aspects of taxonomy. All developers of attack detection systems and organizations that use CADS should understand and study their classification in order to choose the best solutions for information security systems. In the study of various aspects of taxonomy and the application of various options, we can achieve a higher level of security of information systems.

The systems for detecting abnormal behavior are based on the fact that CADS has some features that characterize the correct or permissible behavior of the object of observation. The block diagram of the cyber security of the information system is presented in Fig. 2.

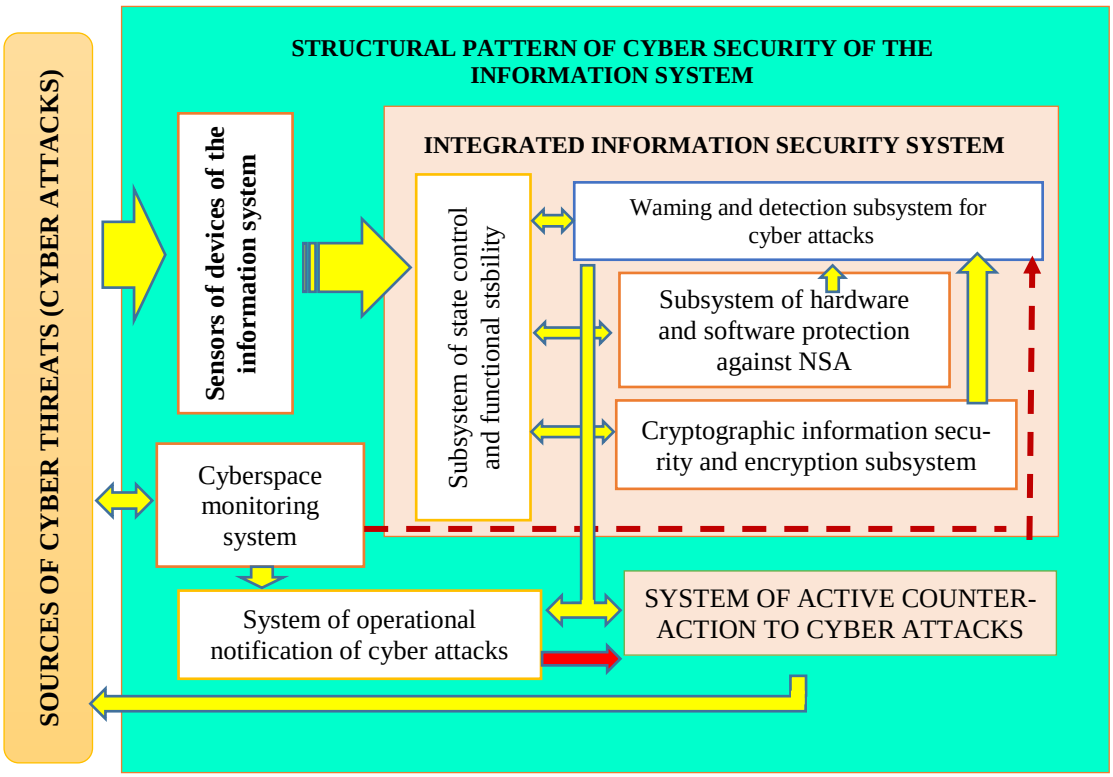

Fig. 2. Structural pattern of the cyber security of the information system

Sensors of cyber intrusion devices identify unusual behavior, anomalies in the operation of a single object. The difficulties of their application in practice are associated with the instability of the objects themselves, which

are protected, and with external objects interacting with them. The object of observation can be the network as a whole, a separate computer, network service, user, etc. Sensors operate on the condition that the intruder violates the normal functioning of the information system.

The measures and methods traditionally used to detect abnormalities include the following:

- threshold values: the observation of an object is expressed in the form of numerical intervals; exceeding these intervals is considered to be an abnormal behavior; thresholds can be static and dynamic;
- statistical measures: the decision on the availability of a cyber attack is taken on the basis of a large number of data collected through their statistical pre-processing;
- parametric: for the detection of a cyber attack a special "normal system profile" is constructed on the basis of templates (some policy which this object must usually follow);
- nonparametric: the profile is built on the basis of observation of the object during the training period;
- measures based on rules (signatures): they are very similar to nonparametric statistical measures; in the period of training an idea of the normal behavior of the object is being formed, which is written in the form of special "rules";
- other measures: neural networks, genetic algorithms, which allow to classify some set of known sensor-indicator signs; in modern CADS the first two methods are mainly used.

Usually, abnormal activity detection systems use logging books and current user activity as a data source for analysis. The *advantages* of cyberattack detection systems based on the technology of detecting abnormal behavior can be estimated as follows:

- anomaly detection systems are capable of detecting new types of cyber attacks, the signatures for which have not yet been developed;
- they do not require renewal of signatures and rules of detection of cyber attacks;
- detection of anomalies generates information that can be used in criminal detection systems.

The *disadvantages* of systems based on the technology of detecting abnormal behavior are:

- systems require long and qualitative training;
- systems generate many mistakes of the second kind;
- systems are usually too slow at work and require a large amount of computing resources.

Let's consider one of the effective methods of detecting intrusions and cyber attacks, which is based on the signature approach. Signatory methods allow you to describe a cyber attack with a set of rules or using a formal model, which can be used as a character string, semantic expression in a special language, etc. The essence of this method is to use a specialized database of templates (signatures) of cyber attacks to find actions which fall under the definition of “cyberattack”.

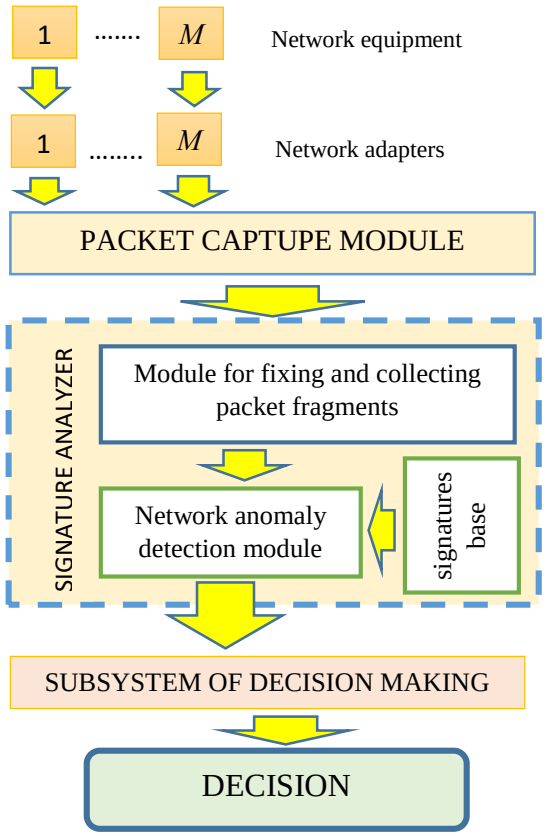

Fig. 3. The universal analyzer structure

The signature method can protect from a viral or hacker cyber attack when its signature is already known (for example, the unchanged fragment of the body of the virus) and it is included in the database of CADS. If the network is experiencing the first attack from the outside, the first infection is still unknown, and the database simply lacks the signature for its search - the signature method CADS will not be able to signal the danger because it considers the attacking activity to be legitimate.

Most of the existing software products which claim to use the signature method, in fact, realize the most primitive way of signature recognition. In such systems, the signature method is implemented as an algorithm that examines only the dynamics of cyberattack development. And it is based on a state machine to assess the scenario of the developing attack. According to the plan, this approach should allow tracking the dynamics of the development of cyber attacks in accordance with the actions of the intruder, while as the module for data collection even the systems for detecting cyber attacks can be used.

Thus, the effectiveness of the signature CADS is determined by three main factors: the efficiency of refinement of the signature base, its completeness from the point of view of the determination of the signature of the cyber attack, as well as the presence of intelligent algorithms for reducing the attacking party's actions to some basic steps, within which there is a comparison with the signatures.

In order to implement the chosen method of determination and identification of CADS, models of the signature and statistical analyzers of network traffic are offered,and the fuzzy intellectual system is used to determine the sources of cyber-media and the choice of solutions for their elimination.

The structure of the universal signature parser flow packets of network traffic is presented in Fig. 3. The mechanism of functioning of the signature analyzer includes two stages: filtering and collecting fragments of packages, recognition of cyber-criminals by signatures.

The work of the analyzer is described by the following model. Denote the network traffic coming from the packet capture module, as a flow in the form of a set $S = \{s_i\}_1^n$, where n is the total number of packets. The base of signatures can be represented as a set of B, which combines signature type clusters $B_j = \{b_{jk}\}_1^K$, $j = \overline{1,m}$:

$$\boldsymbol{B} = \boldsymbol{B}_1 \cup \boldsymbol{B}_2 \cup \ldots \cup \boldsymbol{B}_m = \bigcup_{j=1}^{m} \boldsymbol{B}_j \qquad (1)$$

where m – the number of clusters of signatures;

B_j – j- cluster, which is a set of identical signatures;

K – total number of signatures in the j-cluster.

The input of the response module receives a signal only if $S \subseteq B$.

When developing a statistical analyzer, a model based on the analysis of the mean value and the mean square deviation of the network traffic parameters is proposed. This method is based on comparing the local (current) characteristics Y_b of the flow of packets with averaged over a period of time (global) characteristics y_b. As a statistical characteristic of the flow of packets, a sample average ξ, a sample variance d^2 and a consent criterion χ^2 are used. If the local characteristics are significantly different from global ones, then an abnormal behavior of the packet stream and the likely failure of hardware, software or security policy violations

are concluded. The structure of a statistical analyzer that implements this method of detecting cyber attacks is shown in Fig. 4**.**

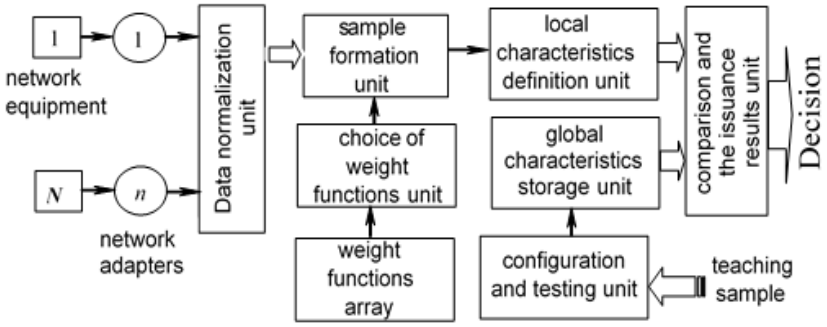

Fig. 4. The statistical analyzer structure

The work of the statistical analyzer is described by the following model. The numeric value $X_i \{x_{\min} \le X_i \le x_{\max}\}$ is a certain event in the flow of network events at a time t_i, $i = \overline{1,n}$. The set of values is characterized by the mean value $\bar{x}$ and variance σ_x of the value *X*. To determine the local characteristics, the average value $\bar{x}$ is calculated not for the whole stream of *N* events, but only for the last *n* events. For this purpose we use the weight function *F(z)* and the local characteristics can be calculated using the following formula:

$$W(N) = \sum_{i=1}^{N} F(t_N - t_i) f(X_i) \quad (2)$$

As a weight function *F(z)* the function of the form for finding *W(N)* was chosen:

$$F_S(z) = \frac{1}{k_S} \sum_{j=1}^{S} \frac{(z/t)^j}{j!} \exp(-z/t) \quad (3)$$

where *t* - is the time interval on which local characteristics are calculated;

k_S - rationing factor.

To determine the local characteristics, the range of possible values *X* is divided into *B* intervals: $[x_{\min}, x_{\max}] \to [x_0, x_1] ... [x_{B-1}, x_B]$ and the hit frequencies in the corresponding intervals are calculated not for the whole stream, but for the *n* most recent events. Local characteristics are calculated by formulas (2) and (3).

When designing an intellectual (expert) system, an obscure model was chosen. This is due to the fact that a significant amount of information on the causes and source of cyber attacks (CA) can only be obtained expertly or in the form of heuristic descriptions of processes. To determine the sources of CA security system should be represented by the model of the information network on which it is oriented. Such a model divides the process of the information moving between computers across the network environment to several levels. Thus, the primary security problem can be represented by the decomposition of security tasks at individual levels of the network.

Represent a separate level of security in the form of a nonlinear object with a plurality of input variables $\{x_i\}, i = \overline{1,n}$ and one output variable *y*:

$$y = f_y(x_1, x_2, ..., x_n) \quad (4)$$

As input variables, we will select signs of CA sources. The output variable *y* is a network status indicator. The model uses the following assumptions and limitati:

- input variables $\{x_i\}$ within one level are independent;
- separate network functions are isolated on each of the network levels.

Integrated Intelligent Decision-making Support System (IIDmSS) for identifying intruders contains a set of functional components which allow you to automate control actions as much as possible when changing the security situation. The structure of the decision-making information system for determining cyber intrusions is presented in Fig. 5.

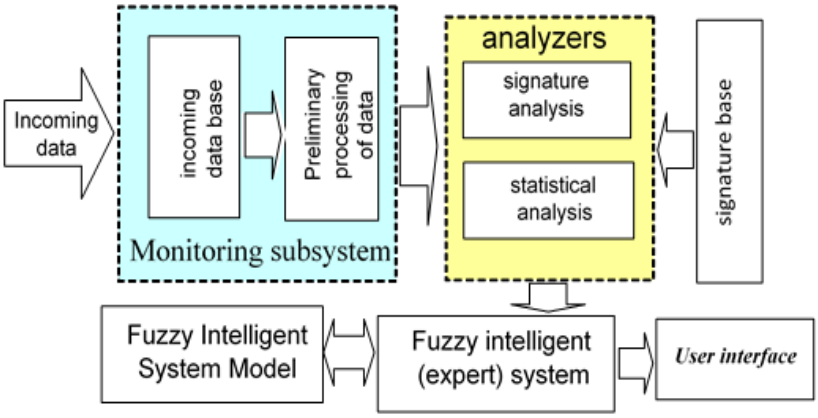

Fig. 5. The structure of the decision-making information system

The current state of cyber attack detection systems for the information systems is full of disadvantages and vulnerabilities, which, unfortunately, allow harmful influences to damage information security systems. The transition from the search for cyber attack signatures to the identification of threats to information security should contribute to radically change this situation by reducing the backlog distance in the development of information security systems from cyber attack systems.

Reference

1. Buhaiskyi, K. V. *Problems of building information security systems* // "Information Security/ Informatsionnaya bezopasnost". – M.: BHV, 2008. – 250 p.

2. Masyuk, M. I., *UAA: theory and practice* // "Special equipment ". – K.: Ruta, 2003. – 300 p.

3. Malyuk, A. A., Pazizin, S. V., Pogozhin, N. S. *Introduction to Information Security in Automated Systems*. – M.: Goryachaya liniya–Telekom, 2001. – 148 p.

4. Domarev, V. V. Security of information technology. Methodology for creating protection systems. – Kiev: OOO «TID DS», 2001. – 688 p.

5. Debar, H., Dacier, M., and Wespi, A. (1999), "Towards a Taxonomy of Intrusion Detection Systems," Computer Networks, vol. 31, 1999, pp. 805 – 22.

6. Debar, H., Dacier, M., and Wespi, A. (2000), "A Revised Taxonomy for Intrusion-Detection Systems," presented at Annales des Télécommunications, vol. 55, 2000, pp. 361 – 78.

7. Kabiri, P., and Ghorbani, A., A. (2005), "Research on Intrusion Detection and Response: A Survey", International Journal of Network Security, Vol.1, No.2, Sep. 2005, pp.84 – 102.

8. Amer, S.H., Hamilton, J.A., "Intrusion Detection Systems, (IDS) Taxonomy – A Short Review," DOD Software Tech News, vol. 13, no. 2, June 2010, DOD Data & Analysis Center for Software, Air Force Research Laboratory, Rome, N.Y., pp. 23 – 30.

9. Ali A. Ghorbani, Wei Lu, and Mahbod Tavallaee, Network Intrusion Detection and Prevention: concepts and techniques. London: Springer, 2010, pp. 27 – 49.

10. Manasi G.; Rana; Yadav, "Taxonomy of Anomaly Based Intrusion Detection System: A Review" International Journal of Scientific and Research Publications, Vol. 2, Issue 12, Dec. 2012.

11. Babenko L. K. Development of a comprehensive system for detecting attacks / L. K. Babenko, O.B. Makarevich, O.Yu. Peskova // I Information security: materials V intern. sci.-pract. conf. 2003. № 4(33), pp. 235 – 239.