RIGA TECHNICAL UNIVERSITY

Faculty of Computer Science and Informatin Technology

Institute of Information Technology

Pāvels OSIPOVS

Doctoral student of program “Information Technology”

THE USE OF PERSONAL ADAPTIVE BEHAVIOR PROFILE FOR DETECTING ANOMALOUS ACTIVITY OF ELECTRONIC INFORMATION SYSTEM USER

Summary of Doctoral Thesis

Scientific supervisor

Dr. habil. sc. comp., professor

A. BORISOVS

RTU Press

Rīga 2015

Osipovs P. The Use Of Personal Adaptive Behavior Profile For Detecting Anomalous Activity Of Electronic Information System User. Summary of Doctoral Thesis.— R.: RTU, 2015.— 36 lpp.

Printed according to the decision of the RTU Institute of Information Technology Board meeting, July 3, 2015, Protocol No. 12100-4.1/4

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».

ISBN 978-9934-10-680-4

DOCTORAL THESIS
IS SUBMITTED FOR THE DOCTOR'S DEGREE IN
ENGINEERING SCIENCE AT RIGA TECHNICAL UNIVERSITY

The defence of the thesis submitted for the doctoral degree in engineering science (Technology of Riga Technical University, in 1/3 Meza Street, auditorium 202, at 14:30, on Juny 1, 2015.

OFFICIAL REVIEWERS

Professor, Dr. habil. sc. comp. Valērijs Zagurskis
Riga Tehnical university, Latvia

Dr. sc. ing. Mihails Savrasovs
Transport and Telecommunication institute, Latvia

Professor, Dr. tech. sc. Aleksandr Koļesņikov
Immanuel Kant Baltic Federal University, Russia

DECLARATION

I hereby confirm that I have developed this thesis submitted for the doctoral degree at Riga Technical University. This thesis has not been submitted for the doctoral degree at any other university.

Pāvels Osipovs ……. (*signature*)

Date: …….

The doctoral thesis is written in Latvian and includes introduction, 5 sections, result analysis and conclusions, 21 tables, 91 figures, overall it consists of 144 pages. The bibliography contains 108 references.

CONTENTS

INTRODUCTION

This research investigates the problem of calculating the level of the anomalous behavior of the user of an electronic information system. This type of problem is important for a variety of electronic systems that provide access to financial, medical, military and similar sensitive data. Such systems are well protected from external threats; however, nowadays the biggest danger comes from internal threats which are caused by lack of monitoring activity of already authorized users. At the same time, financial, reputation and information losses due to loss of important information in large companies can be significant [18].

Theme topicality

The detection of anomalous behavior of electronic information system user requires complex method for its solving [71]. Existing decisions focus on monitoring anomalies of separate components of the target system. Creating a complex approach that provides protection at all levels of functioning of complex information system is a difficult theoretical and technical problem. The base approach described in this research is based on the use of Markov chains [35] and allows obtaining effective and fast (by using the Markov property [50, 62] of the lack of memory) classifier of the user request level of the anomaly. Further increase of its calculation rate is topical because requirements to user queries processing time constantly increase. Methods for improving the efficiency of basic approach are also developed.

Research objective

The main objective of this research is to develop and study possibilities how to increase efficiency of the user's anomaly behavior detection, to the usage of behavior profile adjusted to specifics of the appropriate person behavior and taking into consideration initial set of requirements.

To achieve the set aim, taking into account the existing limitations, it is necessary to implement the following tasks:

- to investigate the existing approaches to formalization of user behavior and their use in anomalous activity detection;
- to investigate possibilities of realization of User Behavior Profile and Personal Adaptive Profile of User Behavior;
- to develop methodology for comparing User Behavior Profile and Personal Adaptive Profile of User Behavior and using it to implement a comparative analysis;
- to develop experimental software environment and to conduct experiments for comparing efficiency of anomalous behavior detection using both approaches. If

possible, to show that Personal Adaptive Profile of User Behavior in general will be the best classifier of anomalous behavior;

- to develop methodology for introduction of the developed approach to modern information systems of different structure.

Object of research

The object of the research is a user behavior profile based on Markov chains. This approach is used to determine the level of anomality in behavior of the the user of complex electronic information systems.

Hypothesis of research

Basic hypothesis set in this thesis is as follows: user behavior profile based only on data about his personal interests within the framework of the system will most effectively calculate the anomaly level in his (user) behavior.

Within the framework of basic hypothesis the following lower-level hypotheses are set:

- the use of the graph of Markov chain is a rapid and effective method of formalization user behavior;
- there is the possibility to compare efficiency of user behavior profiles;
- behavior of different users of the target system leads to different behavior profiles.

Research methods

When building a profile of user behavior, the methods of graph theory and Markov chains are used. The evaluation of the effectiveness of established behavior profiles is carried out by methods of information theory, mathematical statistics, probability theory and frequency analysis. To generate synthetic data on the behavior of users, data mining methods and probability theory is used; for behavior profile training and validation testing machine learning approaches are employed. As a basic research method, a technique of carrying out experiments is established which consists in building an experimental system that allows testing different characteristics of the profiles created.

Scientific novelty

The main achievements described in the dissertation that meet the criteria of scientific novelty, are the following:

- As a more efficient classifier anomalous behavior a method is proposed, that takes into account the behavior of each user of the target information system.
- A method for dynamic changing the threshold for the level of abnormality, is proposed for more accurate operation of the final classifier type of behavior.

- A method for the generation of synthetic data containing different priorities of user's interests is proposed by the use of the exponential distribution with different values of its parameters. Also the possibility of using other types of distributions is pointed out.
- A technique of comparing the effectiveness of behavior profiles is implemented using different approaches trained on data having a variety of characteristics and containing different values of internal variables.
- Obtained estimates of the impact of the internal parameters of the profile behavior on the accuracy of classification type of behavior.

The practical significance

Fundamental difference of the researched type of attacks from other methods of intrusions is that after successful passing of all authorization and authentication procedures, the user has an access to plenty of different services, but specifics of his professional activity increases frequency of usage such services, which are used more frequently. So the task is to detect atypical usage of the system services, in case if they are fundamentally available for usage.

At present the threat of internal attacks is important among various electronic information systems. The considered method can be realized in various information systems:

- different public and private institutions [medical institutions, credit organizations, e-libraries, archives, local governments, political and public organizations, police, military, tax and bank authorities] have access to the information that has to be limited;
- it is possible to create profile of typical owner of portable devices, for example — mobile phones. In this case, features of usage of device will be analogue of person's digital signature;
- the considered method can be used not only for the construction of behavior profile, but also in other tasks, where sets of data determine the specific of their author. For example, construction of author's writing style based on his works;
- the approach developed for the generation of artificial data about the use of the system by a person with various priorities of interests can also be used for development of different software systems.

Approbation of the work

The main stages of research and their results are presented at the following scientific conferences:

1. DMC 2013 — prudsys User Days 2013, Jule 2013, Berlin. www.data-mining-cup.de/en/ *(Osipov P. Anomaly Detection Using User Behavior Profile and Its Implementation into a Distributed Information System).*

2. IADIS Multi Conference on Computer Science and Information Systems 2013 (MCCSIS 2013), July 2013, Prague. www.iadisportal.org/mccsis 2013 (*Osipov P. User Behavior Profile Implementation for Distributed Information System*).
3. XVI International Youth Forum „Radio Electronics and Youth in XXI century", 17–19 April 2012, Kharkov, Ukraine *(Osipov P. A. Identification of Transaction Types using standard Clinical Document Architecture).*
4. 1st International Symposium "Hybrid and Synergies Intelligent Systems: Theory and Practice", (GISIS 2012), 29 June – 2 July 2012, Kaliningrad, Russia *(Osipovs P., Borisovs A. Approaches to the analysis of information systems user behavior modeling).*
5. RTU 53rd International Scientific Conference dedicated to the 150th anniversary and the 1st Congress of World Engineers and Riga Polytechnical Institute/RTU Alumni, 11/12 October 2012, Riga *(Osipovs P., Borisovs A. Modern Approaches to Creating User Behavior Models).*
6. 8th International Scientific School "Modeling and Analysis of Safety and Risk in Complex Systems" (MA SR — 2011), 28 June – 02 July 2011, Saint-Peterburg, Russia. *(Osipov P. A., Borisov A. N. eHealth System Anomaly Activity Detection, Based on User Behavior Model).*
7. RTU 52rd International Scientific Conference. 13–16 October 2011, Riga, Latvia *(Osipovs P., Borisovs A. Deferred — A New Approach to Time-Critical Task Realisation).*
8. Baltic Congress on Future Internet and Communications (BCFIC Riga 2011), 15–18 February 2011, Riga, Latvia. *(Osipov P., Borisov A. Simulation of Typical Behavior User Using Markov Models).*
9. International Scientific Conference of WEB-designers (WebConf 2010), May 2010, Riga, Latvija. *(Osipovs P. Detection of Authorised Users Anomaly Behavior).*
10. RTU 51st International Scientific Conference. Subsection "Information Technology and Management Science". Riga, Latvia. 2010. *(Osipovs P., Borisovs A. Improvement of Markov models for Anomaly Detection Systems).*
11. RTU 50th International Scientific Conference. Riga, Latvija. 2009 *(Osipovs P., Borisovs A. Usage of Ontologies in Systems of Data Exchange).*

Research results obtained in the thesis, are approved in the project: "e-StepControl, identifying suspicious activities", SIA "ABC SOFTWARE", 10.2013–04.2015.

Publications

According to the research carried out in this dissertation, 14 scientific articles was published. Most of the publications are cited in various international digital libraries. Full list

of publications of the author is included in the general list of information sources, presented at the end of this paper.

The main results

The basic results obtained within the framework of research and development of the doctoral thesis can be set as follows.

1. The analysis of the existing modern approaches for anomalous behavior detection of users of information systems is made. During research it has been concluded that none of the considered approaches provide complete implementation of all set requirements.
2. The existing methods of construction and formalization of user behavior based on the use of different variants of the Bayes networks, ontology engineering, mobile agents are investigated and analyzed.
3. Possibilities of application of Markov chains in the tasks of formalization of user behavior profile are investigated and it is shown that the correctness of the set lower-level hypothesis is confirmed.
4. Basic approach for creation and use of general user behavior profile is developed.
5. To increase efficiency of the basic approach for intrusion and anomalous behavior detection, the method using the Personal Adaptive Profile of User Behavior is developed.
6. The methodology for assessment of the efficiency of behavior profiles is developed; it allows implementing a comparative analysis of the Personal Adaptive Profile of User Behavior and General User Behavior Profile.
7. The developed method of dynamic threshold of the anomaly level allows improving the efficiency of user classification on "*normal*" and "*anomalous*".
8. Using the developed methodology for user anomalous behavior detection, a comparative experimental analysis of the efficiency of the Personal Adaptive Profile of User Behavior and General User Behavior Profile is implemented.
9. Different sets of experiments as software system are developed and realized, which allows estimating different descriptions of behavior profiles.

The structure and content of the work

The work consists of an introduction, five parts, conclusion and bibliography. Volume of work — 144 pages, 91 images and 21 tables. The list of references consists of 108 entries.

In the introduction, main area of research is described, the basic concepts are defined, objectives of the research are formulated, and a list of the main tasks and hypotheses is presented.

Chapter 1 is devoted to the current state in research field of the detection of abnormal behavior. Various approaches to anomalies detection in the data and to building profiles of user behavior are described.

Chapter 2 is devoted to the consideration and formal description of the basic algorithm, the definition of metrics used for the classification, evaluating the total complexity of the algorithm and its strengths and weaknesses discuss. In order to improve the basic algorithm the use of Personal Adaptive Behavioral Profile and Anomaly Evaluation Dynamic Threshold is suggested.

Chapter 3 deals with the theoretical approaches to evaluating the effectiveness profiles of user behavior, based on probabilistic characteristics and methods of information theory.

Chapter 4 is devoted to the description of the developed experimental system. The basic restrictions on the process of calculating the level of abnormality of each user request are formulated. Specification requirements for the experimental system are defined. The general structure of the developed system, the main approaches and technologies used in its implementation are described.

Chapter 5 is devoted to the description of all experiments conducted. As part of the achieving main disseratioon goal, each set of experiments produced describes the details of behavior profiles from different sides.

In conclusion, state the main results and conclusions of the work are discussed.

BRIEF DESCRIPTION OF THE WORK CHAPTERS

Chapter 1. Intrusion and anomaly detection task in electronic information systems

Within the framework of this research this term means behavior that differs from normal user behavior. In a general sense "*anomaly*" is deviation from a norm, some error, difference from typical regularity (target for the system). For detection of anomalies, at first we should choose a set of target object parameters used for the calculation of its characteristic values. Then, having a set of measured characteristics it is possible to estimate target object based on their values depending on a set of selected characteristics, as well as on how they describe target object. As a result, the same object can be "*normal*" or "*anomalous*", using the same method, but with different base characteristics.

Within the framework of providing information safety, the concept "*intrusion*" is a series of actions made by a computer or computer network in order to get illegitimate access to the stored information [9]. Intrusion can be made from inside or outside of the target system. Basic types of data threats in case of successful intrusion are the following.

- **Disclosure threat (confidentiality)** — arises in case if confidential information becomes available to the third persons.
- **Integrity threat (integrity)** — arises in case of intentional modification of information.
- **Service refuses threat (availability)** — impossibility for legitimate user to get the required information. Depending on non-availability time of target documents, they can save and lose the status of importance.

The system is considered compromised in case of arising significant probability even one of the mentioned threats.

There are [9] three basic types of intrusion detection techniques.

- **Misuse Detection** (MD) — based systems use the bases of templates of typical attacks that they can detect. Anomaly Detection techniques theoretically allow detect new types of intrusion, based on used models.
- **Anomaly Detection** (AD) approach can detect new types of intrusions, based on its models use.
- **Hybrid methods of IDS** realization (HIDS) can include advantages and disadvantages of Misuse and Anomaly based methods. Mostly systems, realizing such ideology can find out typical types of attacks, as well as new, not known.

In this research, an approach based on the idea of Anomaly Detection is used, so the use of templates of the attacks cannot respond quickly to sudden abnormal behavior of the malicious user, if this attack template is not represented in the existing templates base.

The main requirements to the system for the level anomaly calculation are as follows:

- possibility of system self-education, without the involvement of experts;
- minimum amount of information from the target system is required for analysis;
- maximum level of isolation between the created system and the target system;
- calculation speed of the level of request abnormality must allow doing it in real time.

The analysis of existing approaches functionality has showed that, within the existing requirements, none of the approaches analyzed can cover the required functionality adequately.

Research of the methods [80, 82], used for construction of formal user behavior profiles provided possibility for development of own type of behavior profile, taking into account all required functionality. The analysis of the approaches used for solving anomaly detection task [79], provided the basis for the choice of the most appropriate variant of realization its own system in this area.

Chapter 2. Markov chains in the task of detection anomalous activity

The basic approach used is based on a graph of the Markov chain usage to represent the data regarding the behavior of the user [76, 74, 71]. The overall process of creating a profile behavior, when on the basis of data on the behavior of the user a graph of the Markov chain builds, is shown in Fig. 1.

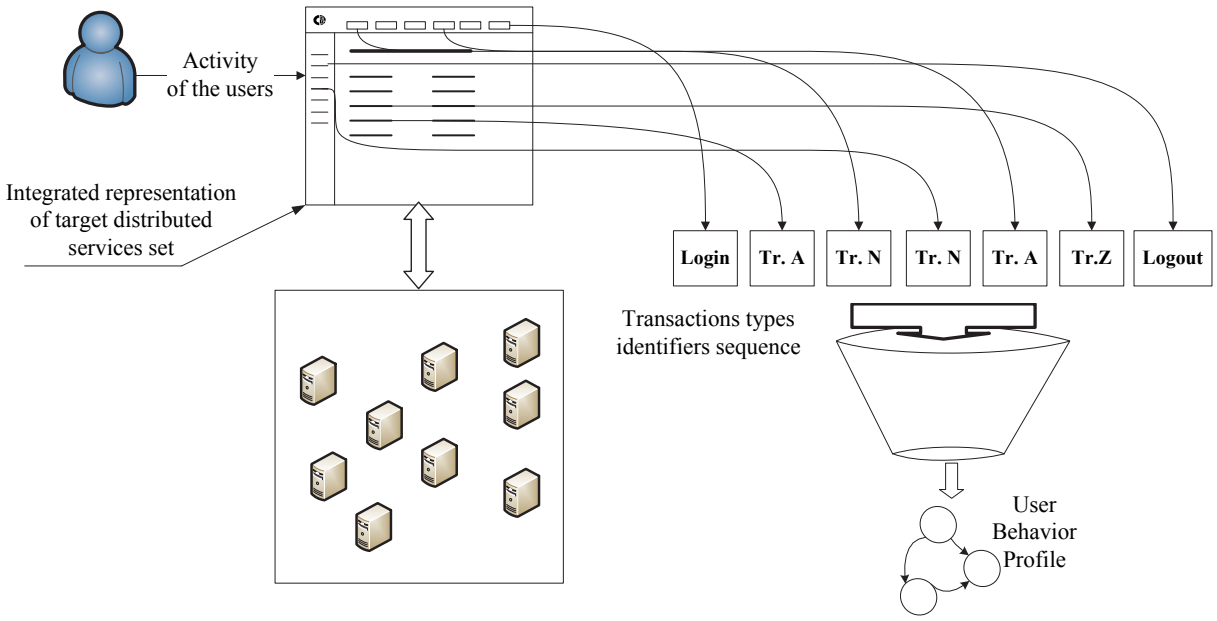

Fig. 1 General approach for UBP creation

Further, already trained profile used to analysis the level of anomality of the user behavior. More details of this process are described on Fig. 2.

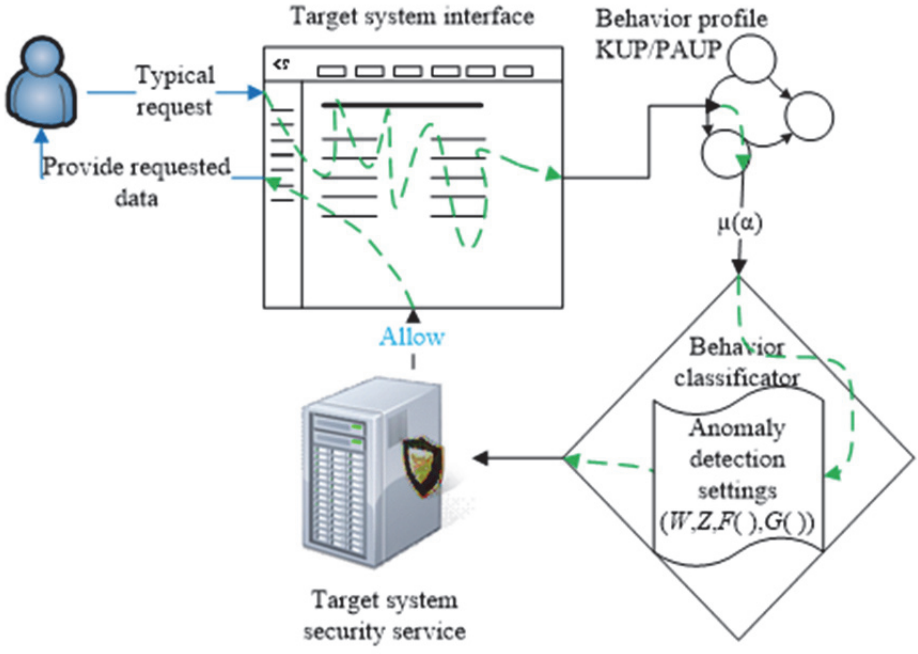

Fig. 2 General scheme of the used approach, in case of "*typical*" behavior of the user

The user inquires a certain transaction at target system. Internal module of security system inquires level of such query anomaly at the module of anomalous behavior detection. The module of calculation of anomaly inquires UBP that corresponds to current user, applying the functions of anomaly layer metrics receipt, the module receives necessary value. If anomaly level is higher than a certain threshold, anomalous behavior is considered as detected, if lower, a query is considered as typical for the user. Final conclusion is send to basic module of security system, which decides how to take into consideration the obtained result.

Formal description of the basic algorithm. Basic algorithm for detecting abnormal activity includes the following steps:

- building the profile of user behavior;
- calculation of metric values and determine the value of the classifier level of abnormality;
- calculation of the level of abnormality of each user's request;
- update of Behavior Profile on the basis of new data of system usage by user.

***Profile education*.** Assume the following notations:

- *Alphabet* Σ — full set of all possible transactions of the user;
- *Window size (w)* — number, amount of alphabet Σ elements; their set forms one node of Markov chain;
- *Z* — the maximum level of penalty in case of absence node in behavior graph what describes the requested state;
- *r* — level of the metric value, for the recognition query as "*abnormal*".

Special empty symbol Ø is added to the alphabet Σ. Initial value of window (*w*) is set. Initial state of Markov chain is determined as suite of length *w*, consisting of zero symbols.

Two operations for traces are set: 1) operation *next* (σ) returns the first symbol of suite σ and moves σ on one position to the left, i.e. *next*(«*abcd*») returns and renews suite to «*bcd*»; 2) operation *shift(σ, x)* moves suite σ to the left and adds the symbol *x* at the end of suite, i.e. *shift*(«*aba*», *c*) = «*bac*».

The cyclic process of Markov chain forming consists of the following steps:

1. suppose *c* = *next(*σ);
2. set <next state> = **shift**(<current state>,c);
3. increase meters for the state <*current state*> and transition (<*current state*>, <*next state*>);
4. renew <*current state*> to the value <*next state*>.

Also sets or updates the value of the transition probability for an arc between nodes describing the previous and current state.

The calculation of the anomaly level for a user's transaction. Implementation of the procedure of calculation of anomaly layer is performed for each transaction inquired by the user. Two internal variables *X* and *Y* are set; their values are tracked during work session. Initially (for the first transaction — *"logging in"*) they have the fixed value; it is allowed to use values *1*, *0* or to calculate on the basis of values of different internal parameters.

On the basis of the constructed Markov chain containing the template of "*normal*" behavior for each step, metrics μ(α) is calculated that determines status of current state and variables *X* and *Y*.

Two variants of *X* and *Y* values calculation are possible at each step, depending on the presence in current behavior graph arc of the transition from the previous to the current state $\beta_i \rightarrow \beta_{i+1}$.

There is a transition:
$$Y = Y + F(s, (s, s'))$$
$$X = X + G(s, (s, s').$$

There is no transition:
$$Y = Y + Z$$
$$X = X + 1.$$

Value of the metrics μ(α) is calculated as the ratio *Y/X*. Metrics μ(α) shows, how current profile predicts suite α, i.e. the less is value, the more precisely Markov chain predicts suite α. As μ parametrized by functions *F*, *G* and number *Z*, then different selection of *F* and *G* will change behavior of classifier that adds possibility to be adjusted according to subject domain.

The classifier *f* constructed from metrics μ as follows. Suite α is classified as anomalous, if metrics μ for current transaction exceeds target threshold value *r*.

Classifier *f*, allowing to determine whether the user's behavior is „*normal*" or „*abnormal*" is determined based on the value of the metric μ by the formula (1), where the value of „1" corresponds to the anomalous and „0" — normal user behavior:

$$f(a) = \begin{cases} 1, & \mu(a) > r \\ 0, & otherwise \end{cases}. \qquad (1)$$

There are several methods for computing functions *X* and *Y* [71]. In this research, the following approaches: *frequency* metrics, *probability metrics* and *metrics based on the value of the local entropy*. The most suitable method is selected depending on the characteristics of the current problem domain. Also *f*(*a*) may be calculated using metrics in same time, and it will be enough to detecting the presence of abnormality in at least one of the results.

Parameters of algorithm and their impact on the quality of result. Since many parameters are used in the construction of classifier, determination of their values impact on the quality of resulting classifier is essential. In addition, it is important to determine the values of metrics of classifier processing power, which are calculated on the basis of both initial sets of traces T and T_{anomal}.

The quality of the created classifier depends on initial training set $T_{traning}$ and the following parameters: size of window w, type of functions $F(s,(s,s`))$ and $G(s,(s,s`))$, value of parameter Z, threshold r. Possible variants of calculation F, G and Z are considered above; w and r are considered below.

Selection of size of window w may be disputable, because small values contain not enough information, but large value is adjusted exactly to training set (*over fitting* [41] — a classic problem of re-training). If to perceive metrics (σ) as measure of difference between Markov chain and analyzed step, i.e. the less it is, the better fragmentation, then discrepancy for all trace in all training set $D(T_{tt})$ can be obtained using the formula: $\frac{\sum \sigma \in T_{tt} \mu(\sigma)}{\sum \sigma \in T_{tt} |\sigma|}$.

It is equal to relation of sum of metrics values for all steps to the sum of the values applied to each node of Markov chain.

Similarly it is calculated discrepancy of trace to anomalous $D(T_{an})$. Size of window w is increased, while difference $D(T_{an}) - D(T_{tt})$ is higher than specified value, i.e. the classifier divides these two subsets.

Regarding to parameter r that is a level, in which the value of metrics for a step is determined as anomalous, in this case it is important to use optimal value. In this research its value is specified so that the number of trigger T_{an} approximately corresponded to training set with the templates of anomalous behavior.

Advantages and disadvantages of the base algorithm. The list of basic advantages and lacks is shown in Table 1. In spite of the fact that the number of lacks is more than the number of advantages, advantages are more significant, especially taking into account the existing limitations and requirements to the system.

Improvement of the base approach. To improve the level of the anomaly detection and to eliminate basic lacks of basic approach, it was suggested to change general profile of class of users with personal profile.

Personal adaptive profile of user behavior. One of basic disadvantages of profile for role of users is that behavior of plenty of users (Fig. 3), even by same role inside of target system can have large uncertainty. As a result, UBP constructed on the basis of such data will describe behavior of some general representative of role. Such general profile describes behavior of each user of role similarly.

Table 1

Advantages and lacks of base approach

Advantages	*Lacks*
1. Possibility to update behavior profile automatically in case of new data about the actions of user. 2. The structure of algorithm allows easy implementation of processing large number of profiles at the same time. It gives an opportunity without considerable effort to increase the system processing power, if it is required increasing the amount of served users. 3. Base algorithm does not require a lot of input data for its work, minimally sufficient evidence of three types: • unique identifier of user; • identifier of its role; • identifier of made transaction. 4. Base algorithm can be easily extended with new types of metrics.	1. Construction of UBP — is difficult and time-consuming task. 2. In the task of updating UBP a significant problem is determination of criteria for activation of user profile updating. 3. Data about behavior of different users united by same role inside of target system have different statistical characteristics; it makes a final model too general. 4. Determination of limit values for UBP, which exceeding is considered as anomalous behavior (intrusion), depends on the specific of current subject domain and cannot be formulated in general. 5. Possibility to hide illegal actions, combining "*good*" and "*bad*" requests. 6. Additional load on the system, arising out of functioning the system of analysis each transaction in real time. 7. Presence of confidential information in data of the system.

To increase accuracy of analysis of user behavior it is assumed to use Personal Adaptive UBP (PAUBP). Such profile will include information about a specific and features of behavior of one — target user (Fig. 4).

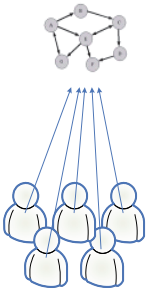

Fig. 3 General profile

Fig. 4 Personal profile

Profile adaptivity is provided by the procedure of **continuous updating of profile** on the basis of user activity. When a user completes session, with the condition that anomalous behavior is not detected, data of session are used for updating the appropriate PAUBP.

As the main advantages of PAUBP approach, more accurate detection of anomalous behavior of a particular user can be distinguished; as well as permanent adjustment to the changes in behavior of the target user and flexible assigning of different profiles to one user.

Dynamic threshold of anomaly level. The method used in the base algorithm, when level of anomaly is determined by fixed value of constant *r*, in general case is not the best. By increase of general value of metrics its level can exceed the set threshold and then even typical behavior will be considered as anomalous. As a result, we get the following situation:

- at the presence of constant "*normal*" behavior, the stationary values of the anomaly metrics is provided;
- then in case of receipt of anomalous queries, the level of metrics decreases;
- then many "*normal*" queries follow, which provide the stationary values of the anomaly metrics again, but already at lower level.

To solve this problem, the method with usage of **dynamic threshold** — r^* is offered. In this case value $r*$ is calculated every time, when a response returned by the module is calculated. In other words, *r** — is the appropriate value of current dynamic threshold for every calculation of anomaly level for received query.

The dynamics of change in the threshold of the anomaly, for behavior of some test user, is shown in Fig. 5.

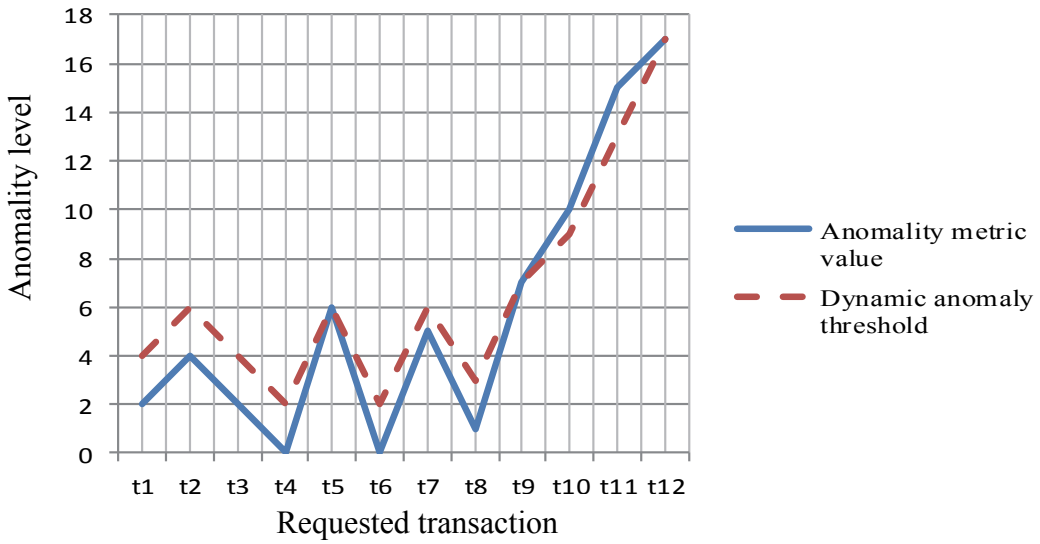

Fig. 5 Levels of the metric values stabilization

Chapter 3. Estimation of efficiency of user behavior profile

There are different methods for estimation of efficiency of information systems [98, 19]. Depending on the specific of the researched task, general efficiency indexes (*reliability*, *authenticity* and *safety*) or set of personal indexes (characterizing *pragmatic*, *technical*,

technological and *operating efficiency* of the system) are used. Depending on the selected basic criterion and the system of indexes, different mathematical methods for estimation of efficiency of difficult systems are used.

Probabilistic characteristics for estimation of quality of user behavior profile. Using probabilistic method, the estimated object is considered as a "*black box*" [29]. Comparison of objects is implemented by implication, by comparison of statistical parameters.

As applied to UBP, such method can be the following — to compare characteristics of metrics issued by the profiles trained on different selections. Significant difference in training selection has to be shown on resulting quality of detection anomalous behavior.

Using such method, profiles can be considered from the point of view of their difference from general profile of class, as well as of statistical properties, when dispersions and values of standard deviation of sets of values of metrics from different profiles are compared.

Interest levels are schematically shown in Fig. 6. *A* in this case is a full set of all possible types of queries in target system. Then *A*' is subset of interests only part of users united by same role inside of target system; *A*'' is a subset *A*' that shows specific interests of a certain user. It is obvious that a profile that shows interests of group *A*', will have other statistical characteristics in comparison with *A*'.

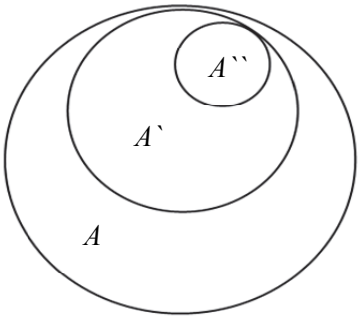

Fig. 6 Interest levels nesting

Presenting gradation of interests in such way, it is possible to determine **probabilistic characteristics for estimation of quality of UBP**. In this case quality of UBP can be estimated by expression "1 — *probability of erroneous conclusion of profile*" [50]. Accordingly, for certain behavior such profile is better, which has less probability of error of both types (when "*anomalous*" behavior is interpreted as "*normal*", or vice versa):

$$Q(\text{UBP}) = 1 - (p(Err_{fpr}) + p(Err_{fnr})),$$

where Q — general concept of profile suitability;
$p(Err_{fpr})$ — probability of the first type error (*False positive error rate*);
$p(Err_{fnr})$ — probability of the second type error (*False negative error rate*).

Usage of such method gives an opportunity to show efficiency of GUBP in comparison with PAUBP, when framework of general profile allow to implement a lot of various actions that will be perceived as "*normal*", while personal profile describes more limited subset, and probability to access for off-site person is less.

Methods of information theory for behaviour profile quality estimation. Basic characteristic for estimation, using information criteria, is *information carrying capacity* [19, 90, 4, 34]. The higher it is, the more effectively the system operates, the more information it receives from every report. So, it is enough to compare efficiency indexes for detection anomalous activity using GUBP and PAUBP. In this case, unit of measure of estimation is entropy of every report on the action implemented by the user. That gives an opportunity to compare these indexes using different methods (Fig. 7).

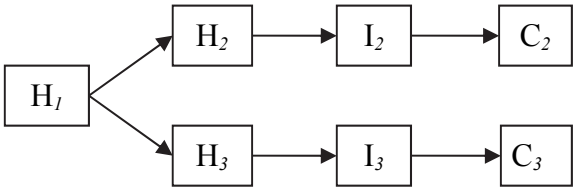

Fig. 7 Efficiency comparison using entropy difference

The figure used the following notation:

H_1 — initial entropy of the system without using UBP;

H_2 — entropy of the system after introduction GUBP;

H_3 — entropy of the system after introduction PAUBP;

I_j — average amount of information, replacing one profile by other ($I_2 = H_1 - H_2$; $I_3 = H_1 - H_3$);

$C_i = I_j \setminus \tau$ — information carrying capacity of the system using the profile *j* for analysis of transaction τ.

As a result, the system with larger value *C* will be more effective accordingly:

$$E = \begin{cases} C_2, & C_2 > C_3 \\ C_3, & \text{otherwise} \end{cases},$$

where E — the most effective profile.

Chapter 4. Programmatic realization of experimental system

In order to show that theoretically operable system can solve real tasks, programmatic platform has been created for the realization of experiments. All operations described in theory are realized using Python [59] and PHP [29], programming languages, the infrastructure for the realization of experiments has been created.

For realization of experimental research of developed theoretical method, programmatic system has been created. General structure of the system is shown in Fig. 8. Basic entry point shows available menu for implementation of experiments. The system is divided into logical modules. Each module implements the part of required functionality.

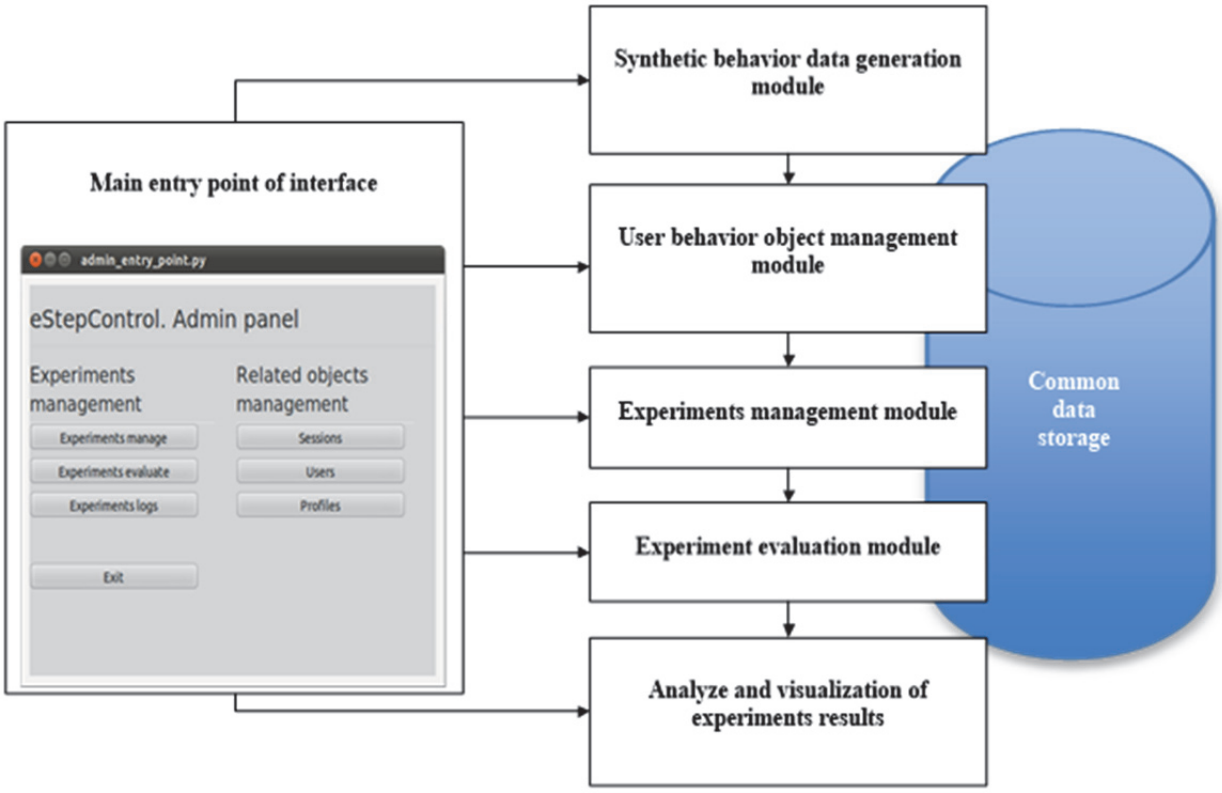

Fig. 8 General strusture of experimantal system

Session generation module users behavior sessions generation with different required characteristics. The generation of behavior sessions can take place in two modes: manual and automatic.

Behavior profile management module creation, training and deleting of UBP programmatic copies.

Experiments management module direct initialization of experiments. Every experiment has a purpose, set of used behavior profiles, and used set of sessions.

Experimental realization module experimental process management. After creation, the experiment can be carried out more than once, its data can be changed, new sessions can be added, and internal parameters can be changed.

Result showing module typical result of the experiment can be dynamics of change of anomaly layer. In this case the module will show the graph. Specialized experiments have other purposes — accordingly get other results. In this case the module can show the graph or to provide raw data for their processing in „*MS Excel*" [10] or "*Statistica*" [44]. Typical

results of the experiment can be dynamics of the change of the value of the anomaly, in this case this module will display its graph.

Limitations of test subject domain. Subject domain has strict requirements to rate of processing of every transaction: the upper threshold of time of processing for 100 models is selected, which is equal to 500 milliseconds.

Basic descriptions of distributed model for realization the system. Practically all servers realized in language Python apply one of following approaches to handle incoming requests:

- usage of system processes;
- usage of light-weight processes (system threads);
- *Deferred* approach.

Deferred approach [92, 81, 34]. The essence of this method is the following: at the receipt of query the module responsible for its processing is called; its function is to process event "*Calculation completed*" and then a server "*forgets*" about the received query and does not consume resources for its support. Later a query is fully processed, a server receives a signal about occurrence of event "*Calculation completed*" and a result of calculation, function that serves a response is called.

In the work, experiments are carried out with all three approaches to handling incoming requests.

Testing of query processing speed. As candidates for the role of basic server processing the incoming requests for the calculation of the anomaly level in the system three popular Python servers have been chosen: Twisted [29], Tornado [102], Cyclone [20].

In Fig. 9 total mean time of response of all servers used in this testing is shown, but not with the emulated artificial delay, but with real loading of the model and implementation of several operations. It is obvious that, as well as in initial tests, *Deferred* approach shows the best results.

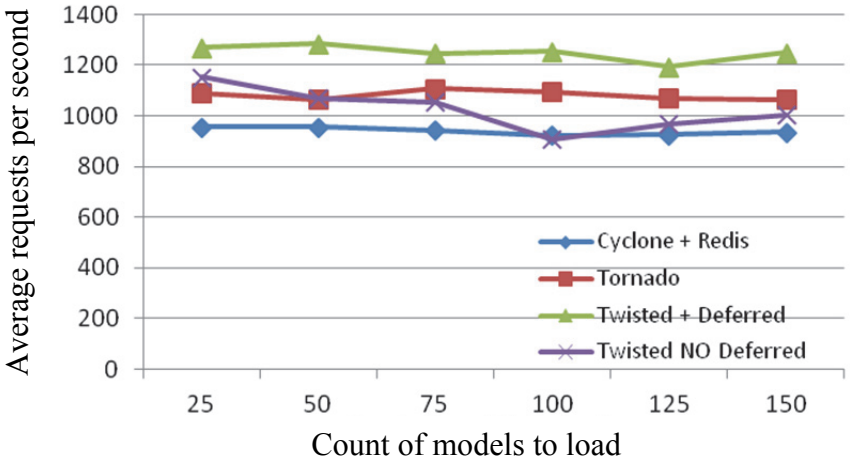

Fig. 9 Response time for different servers

Methodology of the created system introduction. In the terms of real tasks, not always it is possible to understand the efficiency of the developed approach within the framework of different target systems. Approaches to the construction of information systems can differ significantly depending on requirements to their structure, level of availability and data safety. As a result, there are a lot of complex information systems with different architectures, platforms and methods of realization, but it is important for all of them to get possibility to determine the cases of user anomalous behavior.

How can the organizer of complex information system determine that approach for anomalous activity detection developed within the framework of this research can be used?

Basic descriptions of the target information system are specified below to show necessity and validity of introduction the system similar to developed in this thesis.

- Presence of sensitive data. If the system does not use such data, necessity to introduce such complex approach decreases significantly.
- Distributed structure of the target information system realized on the basis of unified integration platform. If the target information system consists of dissimilar blocks not unified in common logical scheme, the possibilities of introduction the created system can be limited.
- Presence of possibility to store and access to data about user queries. If such data are not available, there is no basis for training of behavior profiles.
- Presence of safety providing module in the target information system. The developed approach allows estimating the anomaly level of user queries, but external safety module makes decision about the actions that have to be done.

Introduction of such functionality has to be available within the framework of set time, budget, software and hardware expenses. As a result there is a requirement of unification of external interfaces of the created system in order its introduction in different target platforms will be limited with setting and testing of availability of the required software services.

The created system for the analysis of the anomaly level of user query requires at least three types of the incoming data: *unique identifier of user*, *identifier of his role* and *identifier of transaction he made*. Additionally, for training it is necessary to have data about the previous sessions of work with the system.

As a result the system has to realize at least four internal interconnection interfaces for data transmission of each above mentioned types.

Depending on the availability of all required data, it is possible to specify four types of the target systems (Table 2). Depending on the type of the target system, described in the section, interconnection methods will differ.

Table 2

Structure of public interfaces depending on the type of the target system

Type of the target system structure	***Types of the incoming data***	***Architecture of public interfaces of the system***
Fully available	*Full available*	To use standard interfaces.
Partly available	*Partly available*	For available data to use standard interfaces, for unavailable data realization of specific, non-standard interfaces is required. Aggregating of required data from several sources of the target system is used, or emulation of missing data on the basis of logic maximally suitable for the current target system.
Unavailable	*Unavailable*	All interfaces have to be realized as non-standard scripting interfaces.
Specifically available	*The target system does not provide the presence of all required software or hardware descriptions necessary for functioning of the system*	In case of unavailability of software or hardware resources, the target system administrator will be informed. Then the following actions are possible: • termination of the system work; • attempt to set missing components automatically; • software emulation of the missing nodes; • use of replaced libraries for required software components (for example, use of other NoSQL archive in case of unavailability of Redis).

Determination of necessity criteria for introduction of developed system provided the possibility easily to specify, whether it is necessary to introduce it in the terms of specific of different target information systems. Differentiation of data availability levels in the target information systems allowed to describe the sets of operations required for introduction of developed system at each level.

Chapter 5. Experiments with user behavior profile

Experiments are the main tool of this research. Each set of the conducted experiments considers important part of descriptions. All the experiments carried out are aimed at a common task — to add a comparative description and characteristics of the Common UBP and PUBP, in order to show the advantages of the second approach.

Experiments set № 1. At first, for conducting of experiments, programmatic realization of version 1 has been created. Its basic task was a fundamental estimation of possibility to use the base algorithm for detection anomalous behavior. In the first version possibilities on collection and storage of data test sets, construction of UBP, as well as generation of artificial sets of transactions and calculation of their anomaly level.

A profile is used in the mode of analysis, and the value of metrics for a current step, as well as graph of dynamics of its change for a current session, is specified on Fig. 10. At this stage the system can be in three states:

- ***Initial state* (*warming up*)** — at this stage information in suite of user is not enough for its effective analysis, and the values of metrics can differ significantly;
- ***Normal behavior* —** at this stage the value of metrics has to be in stationary mode;
- ***Anomalous behavior* —** at this stage the value of metrics has to increase continually.

The dynamics of changes in the value of the metric for the first series of experiments corresponds to the theoretically predicted.

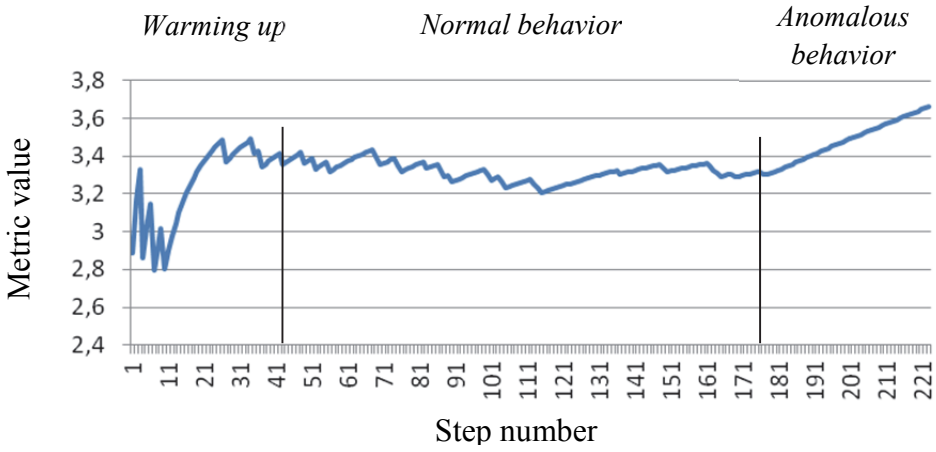

Fig. 10 Metrics behavior at $w = 2$ and $Z = 1,5$

Experiments set № 2. The purpose of this chapter is to formally show that PUUBP will be more effective (in detection anomalous activity) than GUBP. To compare efficiency of two profiles, each of them has been trained on the basis of two different selections created using identical values of internal parameters of algorithm. Each has been used for detection anomalous activity for all possible variants of pairs Profile/data. For example, same_exp determines usage of profile trained on the basis of equally distributed interests for the analysis of set of transactions with exponentially distributed interests (Table 3).

Table 3

Denotation of types of carried out experiments

Profile / Behavior	***Class of users***	***Personal behavior***
Class of users	same_same	same_exp
Personal behavior	exp_same	exp_exp

For example, same_exp determines usage of profile trained on the basis of equally distributed interests for the analysis of set of transactions with exponentially distributed interests.

A summarized graph of all experimental results is shown in Fig. 11. Evidently, that for cases same_same, same_exp and exp_same behavior of profile differs not significantly, and only for the experiment exp_exp difference is significant. Such result confirms a hypothesis that personal profile will be the best classifier for such user on the basis of that behavior he has been trained.

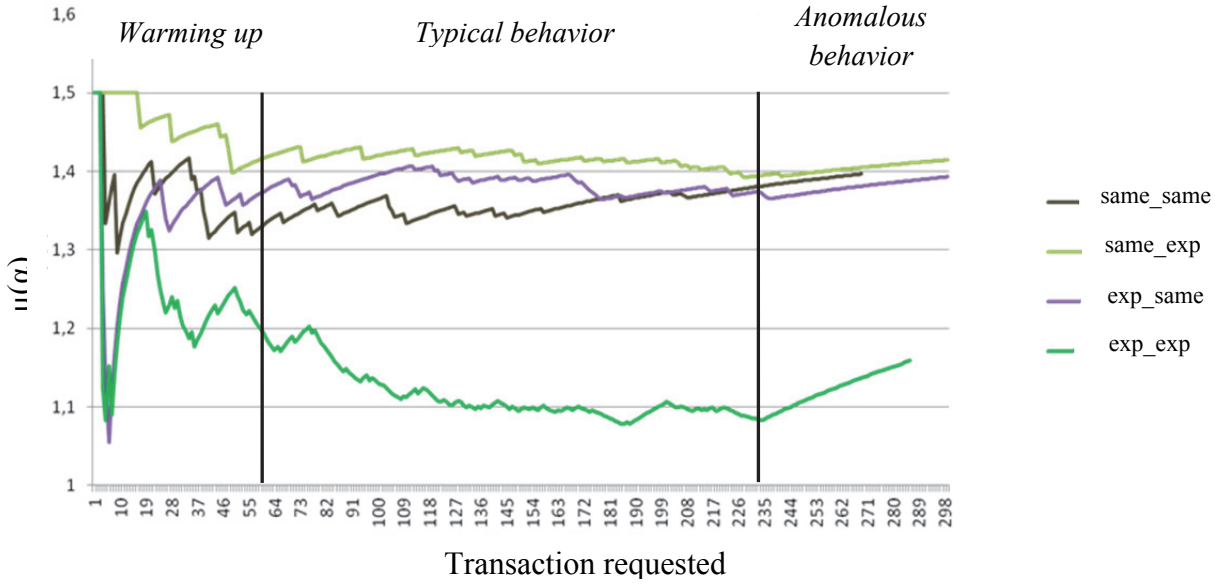

Fig. 11 Summarized graph of all experiments results

Experiments set № 3. The results of the third series of experiments shown on Fig. 12. It is showed that in the course of time the internal weights are well balancing the values of the metrics, thus providing a stationary regime, in case of constant "*normal*" behavior of the user over a large number of requests.

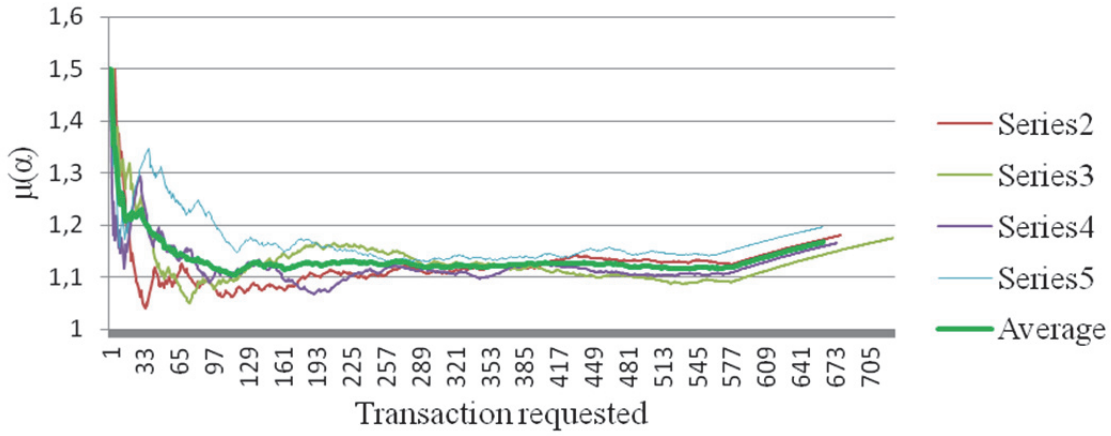

Fig. 12 Results of analysis by profile exp_exp a large amount of "*normal*" activities

Experiments set № 4. Experiments of the fourth series showed a minimal dependence on the specific values (Fig. 13), sets of transactions used for training, because exactly the specifics of user **behavior is important** for a profile, but the **sequence** of queries of transactions is **not so important**.

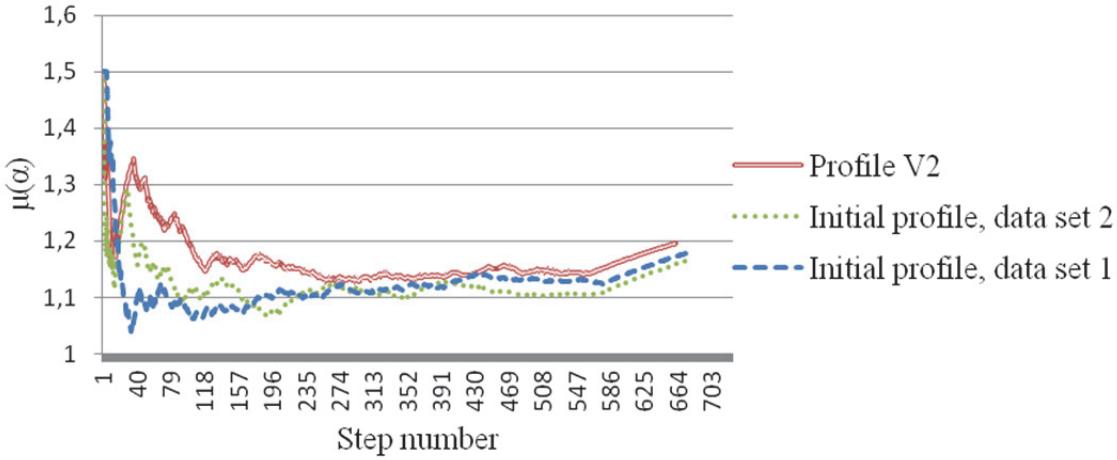

Fig. 13 Two profiles constructed based on the same data

Experiments set № 5. According to the results of the fifth series of experiments we can say that with increasing priorities in the behavior of the user, metrics values also change accordingly. And the greater the difference in priorities, the greater the difference in the values of the level of anomaly (Fig. 14).

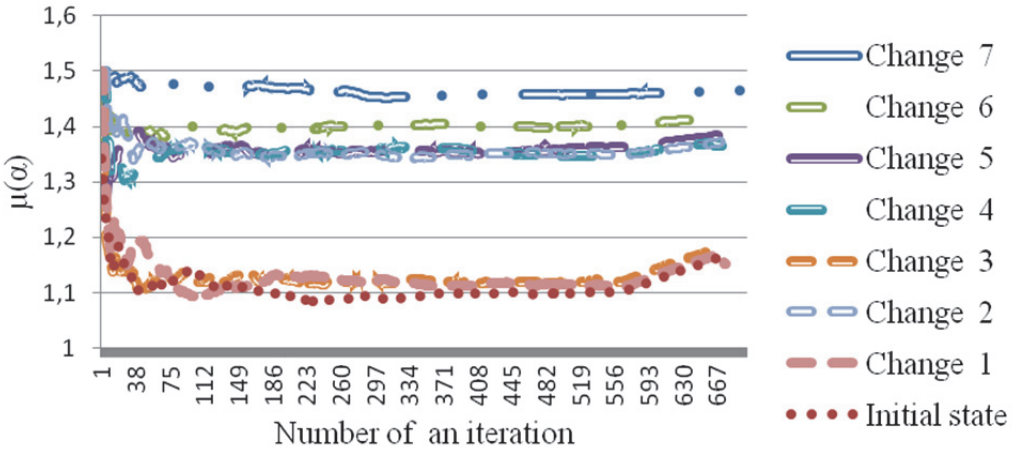

Fig. 14 Summarized values of metrics for the experiment

SUMMARY AND CONSCLUSIONS

Discussion and analysis of results. The main result of this research is development of the methodology of effective determination of the anomaly level in electronic information system user behavior. The efficiency of the created approach complies with initial requirements. The introduction of the Personal Adaptive Profile of User Behavior allowed to increase the accuracy of basic approach for anomalous behavior detection.

To show the efficiency, a series of different experiments representing different descriptions of the created approach have been conducted.

Analyzing the research results, it is possible to assert that the basic hypothesis set at the beginning of the thesis, which is related to the necessity to analyze user personal interests within the framework of the target system, is correct; and Personal Adaptive Profile of User Behavior is the best classifier of user behavior type in comparison with General User Behavior Profile, and more effectively determines the anomaly level of user behavior.

Within the framework of confirmation of basic hypothesis, the lower-level hypotheses set initially have also been confirmed:

Yes — the graph of Markov chain can be used for presentation of information about the features of user behavior;

Yes — there is possibility to compare efficiency of user behavior profiles;

Yes — there is possibility, using the program, to generate data about user behavior, taking into consideration his different interests within the framework of the target system.

The basic results obtained within the framework of research and development of the doctoral thesis can be set as follows.

1) The analysis of the existing modern approaches for anomalous behavior detection of users of information systems has been carried out. During research it has been concluded that none of the considered approaches provide complete implementation of all set requirements.
2) The existing methods of construction and formalization of user behavior based on the use of different variants of the Bayes networks, ontology engineering, mobile agents are investigated and analyzed.
3) Possibilities of application of Markov chains in the tasks of formalization of user behavior profile areinvestigated , which confirmed the correctness of the set lower-level hypothesis.
4) Basic approach for creation and use of general user behavior profile is developed.
5) To increase efficiency of the basic approach for intrusion and anomalous behavior detection, the method using the Personal Adaptive Profile of User Behavior is developed.

6) The methodology for assessment the efficiency of behavior profiles is developed; it allows implementing a comparative analysis of the Personal Adaptive Profile of User Behavior and General User Behavior Profile.
7) The developed method of dynamic threshold of the anomaly level allows improving the efficiency of user classification on "*normal"* and *"anomalous"*.
8) Using the developed methodology for user anomalous behavior detection, a comparative experimental analysis of the efficiency of the Personal Adaptive Profile of User Behavior and General User Behavior Profile is made.

Different sets of experiments as software system are developed and realized, which allow estimating different descriptions of behavior profiles.

The obtained results of the doctoral thesis. The obtained results of the doctoral thesis show the following:

1) On the basis of Markov chains it is possible to quickly create and to effectively use user behavior profiles. The specifics of user behavior uniquely determines the structure of behavior profile.
2) In accordance with the methodology for estimation the efficiency of behavior profiles it is possible to compare the efficiency of different user behavior profiles.
3) Introduction of the Personal Adaptive Profile of User Behavior caused an increase in accuracy of anomalous behavior detection.
4) The results of the conducted experiments showed that the method of the Personal Adaptive Profile of User Behavior is more effective than General User Behavior Profile.

BIBLIOGRAPHY

1. Alavi M., Leidner D. E. Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues // *MIS Quarterly*.— March 2001.— Vol. 25, No. 1.— P. 107–136.— Available on-line on: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8885&rep=rep1&type=pdf. — Last accessed 2014.04.08.
2. Antoniou G., Harmelen F. Web Ontology Language: OWL // Handbook on Ontologies / S. Staab, R. Studer (Eds.).— Berlin: Springer-Verlag, 2004, p. 67–92 (Series: International Handbooks on Information Systems).
3. *ApacheBench*.— Copyright Adam Twiss, Zeus Technology Ltd., 1996.
4. Arndt C. *Information Measures: Information and its Description in Science and Engineering*.— Springer-Verlag, 2013.— ISBN 978-3-540-40855-0. (*Springer Series: Signals and Communication Technology*).

5. Bahder T. B. *Mathematica for Scientists and Engineers*.— Addison-Wesley Publ. Company, 1995.— ISBN-10: 0201540908.

6. Baum L. E., Petrie T. Statistical inference for probabilistic functions of finte state Markov chains // *Annls of Mathematical Statistics*.— 1966 — No. 37.— P. 1554–1563.

7. Beizer B. *Black-Box Testing*.— John Wiley, 1995.— ISBN: 0471120944, 9780471120940.

8. Bernard V. L. *A Guide to Microsoft Excel 2002 for Scientists and Engineers*.— St. Francis Xavier University Nova Scotia, Canada; 2003.— 320 p.— ISBN 0 7506 5613 1.

9. Bhuyan M. H., Bhattacharyya D. K., Kalita J. K. Network Anomaly Detection: Methods, Systems and Tools // *IEEE Communications Surveys & Tutorials*.— 2013.— Vol. 16, Issue 1.— P. 303–336.— ISSN:1553-877X.

10. Billo E. J. *Excel for Scientists and Engineers: Numerical Methods.* 1st ed.— Wiley-Interscience, 2007.— 480 p.— ISBN-13: 978-0471387343.

11. Bishop C. M. Pattern Recognition and Machine Learning (*Information Science and Statistics*).— NY: Springer-Verlag, 2006.

12. Bongard M. *Pattern Recognition*.— SAMS, 2000.— ISBN: 0810491656.

13. Brandes U., Eiglsperger M., Lerner J., Pich C. *Graph Markup Language (GraphML)*.— CRC Press, LLC, 2004.

14. Brooks D. R. *An Introduction to PHP for Scientists and Engineers.*— Springer Science Media, 2008.— ISBN-10: 184800236X.

15. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // *ACM Computing Surveys*.— 2009.— No. 9.— P. 1–72.

16. Chang M., Mathiske B., Smith E., Chaudhuri A., Bebenita M., Gal A., Wimmer Ch., Franz M. *The Impact of Optional Type Information on JIT Compilation Of Dynamically Typed Languages //* 7th Dynamic Languages Symposium (DLS 2011), Portland, Oregon, ACM Press, ISBN 978-1-4503-0939-4, pp. 13–24; October 2011. doi:10.1145/2047849.2047853.

17. Chunfu J., Feng Y. An Intrusion Detection Method Based on Hierarchical Hidden Markov Models // *Wuhan University Jornual of Natural Sciences*.— 2007.— Vol. 12, No. 1.— P. 135–128.

18. *Cost of Cyber Crime Study: United States Benchmark Study of U. S. Companies.*— *Ponemon Institute,* October 2013.— Available on-line on: http://media.scmagazine.com/documents/54/2013_us_ccc_report_final_6-1_13455.pdf. — Last accessed 2014.02.15.

19. Cover T. M., Thomas J. A. *Elements of Information Theory*.— John Wiley & Sons, 1991.— Print ISBN 0-471-06259-6. Online ISBN 0-471-20061-1.

20. *Cyclone.io* project. web-page / Internet.— www.cyclone.io — Last accessed 2014.02.06.

21. *DARPA Agent Markup Language+Ontology Interface Layer / Internet*.— http://www.daml.org/2001/03/daml+oil-index.— Last accessed 2014.10.11.

22. Dasgupta D., Gonsalez F., Yallapu K., Gomez J., Yarramstetti R.. CIDS: An agent-based intrusion detection system // *Computers & Security*.— 2005.— Vol. 24.— P. 387–398.

23. Day J. D., Zimmermann H. The OSI Reference Model // *Proceedings of the IEFJ2*.— 1983.— *Vol. 71, No. 12*.

24. Downey A. B. *How to think like a computer scientist. C++ Version*. 2012 / Internet.— http://www.xplora.org/downloads/Knoppix/books/Open_Book_Project/thinkCScpp.pdf. — Last viewed 2014.09.30.

25. DTI 2002. *Information Security Breaches Survey 2002*. Technical report. Department of Trade & Industry, April 2002. URN 02/318.— Available on-line on: http://www.vicomsoft.com/downloads/learning/dti_security_survey.pdf.— Last accessed 2013.10.12.

26. Duval T., Jouga B., Roger L. The Mitnick Case: How Bayes Could Have Helped // *IFIP International Federation for Information Processing*.— 2005.— Vol. 194/2005.— P. 91–104.— DOI: 10.1007/0-387-31163-7_8.

27. EGEE — *Enabling Grids for E Enabling Grids for E-sciencE* / Internet.— http://www.eu-egee.org/.— Last accessed 2012.04.24.

28. Elliotte R. H., Means W. S. *XML in a Nutshell*. 3rd Edition.— O'Reilly Media; 2009.— Print ISBN: 978-0-596-00764-5.

29. Fettig A. *Twisted Network Programming Essentials*.— O'Reilly Media, 2005.— 238 p.— Print ISBN: 978-0-596-10032-2. ISBN 10: 0-596-10032-9.

30. Fine S, Singer Y, Tishby N. The Hierarchical Hidden Markov Model: Analysis and Applications // *J. Machine Learning*.— 1998.— Vol. 32, No. 1.— P. 41–62.

31. Foster I. *The Grid: Blueprint for a New Computing Infrastructure*.— Morgan Kaufmann Publishers, 1999.— ISBN 1-55860-475-8.

32. *GILDA* virtual laboratory / Internet.— https://gilda.ct.infn.it/.— Last accessed 2012.04.24.

33. Gray J. The Transaction Concept: Virtues and Limitations // *Proceedings of the 7th International Conference on Very Large Databases,* September 9–11, 1981, Cannes, France.— IEEE Press, 1981.— P. 144–154.

34. Gray R. M. *Entropy and Information Theory.* 2nd ed.— New York: Springer US, 2011.— 409 p.— ISBN 978-1-4419-7970-4 (Online).

35. Grinstead C. M., Snell J. L. *Introduction to Probability.* 2nd ed.— American Mathematical Society, 1997.— 510 p.— ISBN-10: 0-8218-9414-5, ISBN-13: 978-0-8218-9414-9.

36. Gunter T. D; Nicolas P. T. The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions // *Journal of Medical Internet Research*.— 2005. Vol. 7, No. 1.— Doi: 10.2196/jmir.7.1.e3. PMC 1550638. PMID 15829475.

37. Hahn B. H., Valentine D. T. *Essential MATLAB for Engineers and Scientists*. 4th edition.— Elsevier Academic Press, 2009.— 416 p.— ISBN: 978-0-12-374883-6.

38. Hahn B. H. *Fortran 90 for Scientists and Engineers*.— London: Butterworth-Heinemann, Cambridge University Press, 1994. ISBN-10: 0-340-60034-9.

39. Harary F. *Graph theory*.— Addison-Wesley Publishing, 1969.

40. Hassanzadeh O. *Introduction to Semantic Web Technologies & Linked Data*, University of Toronto, CS 443: Database Management Systems — Winter 2011.— Available on-line on: http://www.cs.toronto.edu/~oktie/slides/web-of-data-intro.pdf.— Last accessed 2014.04.08.

41. Hawkins D. M. The Problem of Overfitting *// J. Chem. Inf. Comput. Sci*.— 2004.— Vol. 44.— P. 1–12.

42. Haykin S. *Neural Networks: A Comprehensive Foundation*.— New Jersey: Prentice Hall, 1999.

43. *EHR — Electronic Health Records: Manual for Developing Countries*.— WHO Library Cataloguing in Publication Data; World Health Organization, 2006.— ISBN 92 9061 2177.— Available on-line on: http://www.wpro.who.int/publications/docs/EHRmanual.pdf. — Last accessed 2014.05.27.

44. Hill T., Lewicki P. *STATISTICS: Methods and Applications*.— StatSoft. Tulsa. 2007.— Available on-line on: http://www.statsoft.com/textbook/.

45. IMS — *Learner Information Packaging Information Model Specification, Final Specification*. Version 1.0 / Internet.— http://www.imsglobal.org/profiles/lipinfo01.html.— Last accessed 2012.04.24.

46. Isaza G., Castillo A., López M., Castillo L. Towards Ontology-Based Intelligent Model for Intrusion Detection and Prevention // *Advances in Soft Computing*.— 2009.— Vol. 63.— P. 109–116.— DOI: 10.1007/978-3-642-04091-7_14.

47. Ivanov Y, Bobick A. Recognition of Visual Activities and Interactions by Stochastic Parsing // *J. IEEE Trans on Pattern Analysis and Machine Intelligence*.— 2000.— Vol. 22, No. 8.— P. 852–872.

48. Jaiganesh V., Mangayarkarasi S., Dr. Sumathi P. Intrusion Detection Systems: A Survey and Analysis of Classification Techniques // *International Journal of Advanced Research in Computer and Communication Engineering*.— 2013.— Vol. 2, Issue 4.— P. 1629–1635.— ISSN (Print): 2319-5940, ISSN (Online): 2278-1021.

49. Jha S., Kruger L., Kurtx T., Lee Y., Smith A. *A Filtering Approach To Anomaly and Masquerade Detection*. 2005. Technical report, Univ of Wisconsin, Madison.

50. Jha S., Tan K., Maxion R. A. Markov Chains, Classifiers and Intrusion Detection // *Computer Security Foundations Workshop* (CSFW), 2001. Proceedings. 14th IEEE , vol. 1, pp. 206–219, DOI: 10.1109/CSFW.2001.930147. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7408 Last accessed 2014.02.12.

51. Joaquim P. M. *Applied Statistics Using SPSS, STATISTICA, MATLAB and R.* 2nd ed.— Springer Press, 2007. ISBN 978-3-540-71972-4.

52. Johnson N. L., Kotz S., Balakrishnan N. *Continuous univariate distributions*. Vol. 1, 2nd ed.— New York: John Wiley & Sons, 1994.— ISBN 978-0-471-58495-7. (Wiley Series in Probability and Mathematical Statistics: *Applied Probability and Statistics).*

53. Judea P. Causal inference in statistics: An overview // *Statist. Surv.*— 2009.— Vol. 3.— P. 96–146.— Doi:10.1214/09-SS057. http://projecteuclid.org/euclid.ssu/1255440554.

54. Kallenberg O. *Foundations of Modern Probability*, 2nd ed.— Springer-Verlag, 2002.— 650 p. (Springer Series in Statistics).— ISBN 0-387-95313-2.

55. J. Kopena and W. C. Regli, "*DAMLJessKB: A tool for reasoning with the semantic web*," IEEE Intelligent Systems, Vol. 18, pp. 74–77, 2003.

56. Laskey, K. B., Alghamdi, G., Wang, X., Barbara, D., Shackleford, T., Wright, E., Fitzgerald, J. Detecting Threatening Behavior Using Bayesian Networks // *Proceedings of the Conference on Behavioral Representation in Modeling and Simulation*, 2004.

57. Lawson, T. *A Conception of Ontology*.— University of Cambridge, 2004.

58. Liepins G. E. and Vaccaro H. S. Anomaly Detection: Purpose and Framework // *Proceedings of the 12th National Computer Security Conference*, October 1989. P. 495–504.

59. Lucks J. B. *Python — All a Scientist Needs*.— Pycon, 2008.— arXiv:0803.1838v1.

60. M´etivier F. *Scientific Relational Databases using MySQL and Python:* Lecture notes.— Paris: Institut de physique du globe de Paris & Universit´e Paris Diderot, 2014.— 53 p.

61. Manavoglu E., Pavlov D., Lee C. Probabilistic User Behavior Models // *Proceedings of the Third IEEE International Conference on Data Mining* (ICDM'03), November 2003, Melbourne, Florida.— IEEE, 2003, p. 203–210.

62. Markov A. A. *Theory of Algorithms.*— M.: 1954. [Translated by Jacques J. Schorr-Kon and PST staff] Imprint Moscow, Academy of Sciences of the USSR, 1954 [Jerusalem, Israel Program for Scientific Translations, 1961; available from Office of Technical Services, United States Department of Commerce] Added t.p. in Russian Translation of Works of the Mathematical Institute, Academy of Sciences of the USSR, v. 42. Original title: Teoriya algorifmov. [QA248.M2943 Dartmouth College library. U.S. Dept. of Commerce, Office of Technical Services, number OTS 60−51085.].

63. Maxfield B. *Essential MATHCAD for Engineering, Science and Math*.— London: Elsevier Academic Press, 2009.— ISBN: 978-0-12-374783-9.

64. Mea D. V. What is e-Health: The death of telemedicine*? // Journal of Medical Internet Research.— 2001.— Vol. 3, No. 2.*— Doi:10.2196/jmir.3.2.e22.

65. Millman, K. J., Aivazis M. Python for Scientists and Engineers. University of California, Berkeley // *IEEE Computational Science & Engineering.*— 2011. Vol. 13, Issue 2.— P. 9–12.— ISSN: 1521-9615.

66. Mun G. J., Kim Y. M., Kim D. K., Noh B. N. Network Intrusion Detection Using Statistical Probability Distribution // Computational Science and Its Applications — ICCSA 2006.— Springer-Verlag Berlin Heidelberg, 2006, p. 340–348, (Lecture Notes in Computer Science, Vol.3981).

67. *Networkx / Internet*.— http://networkx.lanl.gov/.— Last viewed at 2014-02-07.

68. Noy N. F., McGuinness D. L. *Ontology Development 101: A Guide to Creating Your First Ontology.* KSL Technical Report.— Stanford University, Stanford, CA, 94305, 2006.— Availible on-line: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.— Last viewed at 2015-01-07.

69. Orwant J. *Computer Science & Perl Programming: Best of TPJ.* 1st edition.— O'Reilly Media: 2002.— ISBN-10: 0596003102.

70. Osipov P., Borisov A. Advantages of Deferred Approach for Time-Critical Tasks // *Informatica*.— 2014.— Vol. 25, No. 3.— P. 467–484.— DOI: http://dx.doi.org/10.15388/Informatica.2014.24. Cited in: **ACM, DBLP, EBSCO, SCOPUS, INSPEC, IAOR, Cambridge Scientific Abstracts, Mathematical Reviews, MathSciNet, Science Citation Index Expanded, Web of Science.**

71. Osipov P. A., Borisov A. N. System for anomalous activity detection based on Markov models // *Automatic Control and Computer Sciences.*— 2011.— Vol. 45, No. 2.— P. 46–60. Cited in: **SpringerLink, SCOPUS, Academis OneFile, DBLP, Inspec**.

72. Osipov P. A., Mrochko A. E. and Borisov A. N. Identification of Differences of User Behavior Profiles and User Class Templates // *Automatic Control and Computer Sciences.*— 2014.— Vol. 48, No. 2.— P. 65–79. Cited in: **SpringerLink, SCOPUS, Academic OneFile, DBLP, Inspec**.

73. Osipov P., Rinkevics A., Kuleshova G., Borisov A. Markov chains in the task of author's writing style profile construction // *Scientific Journal of Riga Technical University, Information Technology and Management Science*.— 2014.— Vol.17, RTU, Riga, P. 119–125. Cited in: **EBSCO, Google Scholar, Ulrich's International Periodicals Directory, VINITI**.

74. Osipov P., Borisov A. Simulation of Typical Behavior User using Markov Models // *Proceedings of 2011 Baltic Congress on Future Internet and Communications (BCFIC*

2011), 15–18 February 2011, Riga, Latvia.— Riga: Transport and Telecommunication Institute, 2011.— P. 229–236.

75. Osipov P. A. Borisovs A. Identification of Transaction Types using standard Clinical Document Architecture // Proceedings of XVI International Youth Forum „Radio Electronics and Youth in XXI century", 17–19 April 2012, Kharkov, Ukraine. Vol. 6.— Kharkov: Kharkov National University of Radio Electronics, 2012.— P. 150–151.

76. Osipov P. A., Borisov A. N. eHealth System Anomaly Activity Detection, Based on User Behavior Model // Modeling and Analysis of Safety and Risk in Complex Systems: Proceedings of International Scientific School MA SR — 2011 (Saint-Peterburg, Russia, 28 June – 02 July 2011).— SPb.: SUAI, SPb., 2011.— P. 405–412.

77. Osipovs P., Borisovs A. Approaches to the Construction of Behavioural Patterns of Information System Users // *Scientific Journal of Riga Technical University. Information Technology and Management Science.*— 2012.— Vol. 15.— P. 176–182. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

78. Osipovs P., Borisovs A. Approaches to the Creation of Behavioural Patterns of Information System Users // *Scientific Journal of Riga Technical University. Information Technology and Management Science.*— 2012.— Vol. 15.— P.58–64. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

79. Osipovs P., Borisovs A. Non-Signature-Based Methods for Anomaly Detection // *Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.*— 2010.— Vol. 44.— P. 106–110. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

80. Osipovs P., Borisovs A. Usage of Ontologies in Systems of Data Exchange // *Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.*— 2009.— Vol. 40.— P. 108–116. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

81. Osipovs P., Borisovs A. Use of Deferred approach in Scientific Applications // *Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.*— 2011.— Vol. 49.— P.139–144. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

82. Osipovs, P., Borisovs, A. Practice of Web Data Mining Methods Application // *Scientific Journal of Riga Technical University. Series 5. Computer Science, Information Technology and Management Science.*— 2009.— Vol. 40.— P. 101–107. Cited in: **EBSCO, CSA/ProQuest, VINITI**.

83. Pearl, J. (1985). Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning (UCLA Technical Report CSD-850017). *Proceedings of the 7th Conference of the Cognitive Science Society,* University of California, Irvine, CA. pp. 329–334. Retrieved 2009-05-01.

84. Phyo A. H., Furnell S. M. A Detection-Oriented Classification of Insider IT Misuse // *Proceedings of the 3rd Security Conference, Las Vegas, USA*, 14–15 April 2004.

85. Pinheiro C., Carlos A. R. *Social Network Analysis in Telecommunications*.— John Wiley & Sons, 2011.— 304 p.— ISBN 978-1-118-01094-5.

86. Pokorny J. NoSQL databases: a step to database scalability in web environment // *International Journal of Web Information Systems*.— 2013.— Vol. 9, No. 1.— P. 69–82. — DOI 10.1108/17440081311316398.

87. Prechelt L. An empirical comparison of C, C++, Java, Perl, Python, Rexx, and Tcl for a search/string-processing program. Fakultat fur Informatik. Universitat Karlsruhe. *Technical Report* 2000-5. March 10, 2000.— Available on-line on http://www.inf.fu-berlin.de/inst/ag-se/teaching/V-EMPIR-2014/doc/jccpprtTR.pdf.— Last accessed 2014-05-27.

88. Razmerita L. Modeling Behavior of Users in Adaptive and Semantic-enhanced Information Systems: The role of a User Ontology // *Proceedings of 5th International Conference of Adaptive Hypermedia and Adaptive Web-Based Systems and Authoring of Adaptive and Adaptable Hypermedia workshop,* P. 69–77.— 29 of July – 1 August 2008, Hanover. *Internet*.— http://www.academia.edu/3134137/Modeling_behavior_of_users_in_adaptive_and_semantic-enhanced_information_systems_The_role_of_a_user_ontology
Last accessed 2014.09.08.

89. Redis / Internet.— http://redis.io/topics/faq.— Last accessed 2014.04.01.

90. Reza F. M. *An Introduction to Information Theory*.— New York: Dover Publications, Inc., 1994.— ISBN 0-486-68210-2.

91. RFC7159: *The JavaScript Object Notation (JSON) Data Interchange Format / Internet.*— http://tools.ietf.org/html/rfc7159.— Last accessed 2013.05.03.

92. Samek M. *Practical UML Statecharts in C/C++: Event-Driven Programming for Embedded Systems*.— CRC Press, Newnes 2008.— 728 p.— ISBN-10: 0750687061; ISBN-13: 978-0750687065.

93. Scarfone K., Mell P. *Guide to Intrusion Detection and Prevention Systems (IDPS): Recommendations of the National Institute of Standards and Technology*. NIST Special Publication 800-94.— Computer Security Division, Information Technology Laboratory, *National Institute of Standards and Technology*, 2007.

94. Seung-Hyun K., Kyong H. K., Jong K., Sung-Je H., Sangwan K. Workflow-Based Authorization Service in the Grid // *Journal of Grid Computing*.— 2004.— No. 2.— P. 43–55.

95. Shelestov, S. Skakun, N. Kussul. Agent-based approach to implementing a model of user behavior Grid-systems // Proceedings of Space Research Institute NASU-NSAU «*Informatyka, kibernetyka ta obchyslyuval'na tekhnika*», Issue. 9 (132).— Donetsk: DonNTU, 2008.— P. 8–14.— ISSN: 1996-1588.

96. Sheskin D. *Handbook of Parametric and Nonparametric Statistical Procedures*.— CRC Press, 2004.— P. 54.— ISBN 1584884401.

97. Shingo T., Susumu D., Shinji S. A user-oriented secure file system on the Grid // *The 3rd IEEE/ACM Int. Symp. on Cluster Computing and the Grid* (CCGrid 2003), Conference Report / Oral Presentation. May, 2003.

98. Shirai K. Interest rate risk modeling using extended lognormal distribution with variable volatility // *Stochastic Modeling*.— International Actuarial Association May 2010.— ISBN: 978-0-9813968-2-8.

99. Shneiderman B. The Relationship Between COBOL and Computer Science // *American Federation of Information Processing Societies* 1985.— AFIPS 0164-l 239/85/040348-352$01 .OO/OO.

100. Sriparasa S. S. *JavaScript and JSON Essentials*.— O'Reilly Media, 2013.— ISBN 10: 1-78328-604-0.

101. Tabini M. PHP as a General-Purpose Language // *Linux Journal*.— Aug 18, 2004 / Internet — http://www.linuxjournal.com/article/6627 Last accessed 2014.10.10.

102. *Tornadoweb*. project web-page / Internet.— www.tornadoweb.org.— Last accessed 2014.02.06.

103. Turban E., Aronson J. E., Liang T. P. *Decision Support Systems and Intelligent Systems*. 7th ed.— Prentice Hall, 2005.

104. Undercoffer J., Pinkston J., Joshi A., Finin T. A Target-Centric Ontology for Intrusion Detection // IJCAI-03 *Workshop on Ontologies and Distributed Systems*, MorganKaufmann Publ., P. 47–58.— Acapulco, 9 August 2003 / Internet.— http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.727 Last accessed 2014.10.01

105. Wood R. *C Programming for Scientists and Engineers*.— Penton Press, 2002.— ISBN 1 8571 8030 5.

106. Zinky J., Shapiro R., Siracuse S., Wright T. Experience with Dynamic Crosscutting in Cougaar // *Lecture Notes in Computer Science*.— 2010.— Vol. 4803.— P. 595–612.— DOI: 10.1007/978-3-540-76848-7_41.

107. Осипов П. А., Борисов А. Н. Система обнаружения аномальных действий на основе моделей Маркова // *Автоматика и вычислительная техника.*— 2011.— № 2.— С. 46–60. Cited in: **SpringerLink, Ulric's International Periodicals Directory, VINITI**.

108. Осипов П. А., Мрочко А. Е., Борисов А. Н. Идентификация отличий профиля поведения пользователя и шаблона класса пользователей // *Автоматика и вычислительная техника.*— 2014.— № 2.— С.5–24. Cited in: **SpringerLink, Ulric's International Periodicals Directory, VINITI**.