VERSITA

Information Technology and Management Science

doi: 10.2478/v10313-012-0008-5

2012 /15

Approaches to the Creation of Behavioural Patterns of Information System Users

Pavels Osipovs[1], Arkady Borisov[2], [1-2]*Riga Technical University*

***Abstract* – During the development of the system for anomaly detection in the electronic information system, there is a need to review the existing research in the field of user behaviour modelling. Approaches to user behaviour modelling are very diverse: the algorithms based on neural networks, agent-based approach, Bayesian networks and ontologies. Each approach has its advantages and disadvantages, features, and the applicability for the infrastructure of modern complex electronic systems.**

***Keywords* – anomaly detection, user behaviour model, ontology, Bayesian network, neural network, multi-agent system**

I. Introduction

Nowadays, the task for creating a model of behaviour of information system users is considered in a variety of academic and applied research projects. Various sources [3] indicate that 60 to 93% of the most dangerous attacks on the information infrastructure have been made just using the accounts of authorized users. It is especially important to deal with the threats of this type of systems that operate on sensitive data that can be used by terrorists, criminals, spies or competitors. Such data can be *stolen*, *destroyed* or *compromised*.

In addition to common security protocols used, specifics of some system problems require the analysis of users' activity goals. One possible approach to provide security from internal threats is to use models of user behaviour. In this case, the existing formal model of user behaviour can be compared with other models used to predict future behaviour, to obtain information about the most typical behaviour, and especially to help detect anomalous behaviour.

In the process of developing this type of a system [13] there is a need to review the existing research in the field of modelling behaviour of information system users. By understanding the available approaches and the results of other researchers, it is important to assess the results of target system design.

The paper considers the various modelling methods of behaviour of information system users. Approaches to user behaviour modelling are very diverse, for example, purely mathematical or statistical models are used, as well as a variety of hybrid algorithms. As the foundation of behaviour modelling, it is possible to use Markov models, hidden Markov models, neural networks, Bayesian networks, ontologies, behavioural patterns at the OS level [1], genetic algorithms, and the analysis of traffic at the network level, based on probability and entropy models. Each approach has its own features, advantages and disadvantages, and the applicability in the modern infrastructure of complex electronic systems.

II. Methods

A. Bayesian Networks

Let us recall that a Bayesian network – is a probabilistic model, represented as a directed acyclic graph, and represents a set of variables and their probabilistic dependencies.

In [2], the method for detecting internal threats in the information system is described. The problem is solved by constructing a model of user behaviour based on hybrid Bayesian networks (multi-entity Bayesian networks (MEBNs)). Usual approach is used, when the existing user model, built on the basis of the previous behaviour, is compared with the current operations, and the large difference in rates indicates the presence of abnormal behaviour. To prove the consistency of the approach, the implementation of the method is described.

If the user knows about the type of security used in a protected system, he can start changing his behavioural pattern very slowly that does not reveal much difference from session to session, and only the use of the long history of the system use can detect the fact of constant change in his behaviour. It does not necessarily mean that he wilfully does something bad, but still it should cause the security system to pay more attention to it. Despite the presence of a certain order in the human behaviour, it is highly unpredictable. The behaviour is influenced by many external and internal factors, but within the information system itself it is limited to the number of possible actions that can build a finite model.

B. The Approach of the Research

Simple Bayesian Networks

Bayesian probability theory is a powerful tool for building models in the case of uncertainty. An important advantage of these models is the ability to combine both expert data and experimental results for a long time, thus increasing accuracy. Recently, a combined approach is gaining popularity, when the Bayesian model is used in conjunction with the theory of graphs, which allows building more complex and accurate models based on a large number of cross-hypotheses. Bayesian network is a probabilistic model presented as a directed graph, designed to display high-quality relationships and the local probability distribution for the display of quantitative information on the types and levels of relations between concepts.

Fig. 1 presents part of a simple Bayesian model of user behaviour upon requesting the document. The model is represented as a set of random values within the general hypothesis. For example, the random variable *GlobalIntention* indicates the likelihood of ill will in the request. It, in turn, depends on the value of a random variable *Motive*. In the absence of motivation, it is likely that the user will have no illegal intentions, whereas for a motivated user this probability increases significantly.

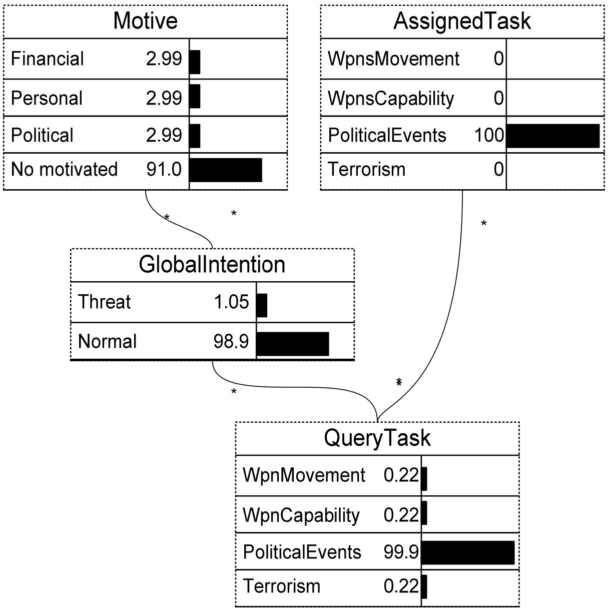

Fig. 1. Part of Bayesian model

Hybrid Bayesian Networks

In standard Bayesian networks, there is a limited range of issues, in which one set of random variables is applied to all problems, and only the symptoms vary from problem to problem. A more flexible representation can be achieved by using only atypical situation as a model.

For this purpose, it is necessary to add the concept of hybrid Bayesian networks, which operate on parameterized structures called MEBN Fragments (*MFrags*). Each modular component is *MFrag* representing fairly small, self-sufficient, and conceptually significant part to support or refute the current hypothesis. *MFrag* enables one to simulate various hypotheses, creating a chain of interrelated concepts. The modular design allows combining and re-using previously created elements in the new research. In general, each *MFrag* encapsulates the functionality of a simple Bayesian model element.

For example, Fig. 2 shows how to use *MFrag* model, to present the model from Fig. 1. Each *MFrag* contains a set of values of random variables (shown in white), the distribution parameters are also included in the model, the input random values (marked in light gray), and their values depend on the internal random variables, random variables and context (marked dark gray), which should be set to *True*, to show the importance of the distribution defined in the resident random variables *MFrag*. All random variables have the input value, called entities. For example, the query *MFrag* shown in Fig. 2 will be correct, when the entity *u* (indicating the user) is the value of *PerformingUser(q)*, for the entity *q* (denoting a request) – i.e. when a particular user performs a specific request.

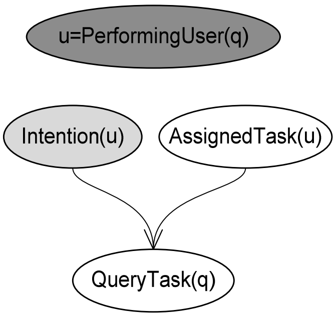

Fig. 2. Example of MFrag dependencies

Local distribution and their parameters used in *MFrag* are extracted from real data. This approach can be used to describe complex models using a large number of participants, instruments and computer systems. The final model is constructed as a Bayesian network consisting of interconnected *MFrag*. The training model is applied in real time to the description of the current user behaviour; it can be used to detect anomalous behaviour, as well as for its study and analysis.

Experimental Model

For the experiments, it has been necessary to set up a test system consisting of seven *MFrags* and involving the modelling of the behaviour of users, which makes requests to the system and works with documents.

The following entities have been created:

User – *MFrag* describes a single user. It describes the identity of the user and his motives and intentions.

User Background – describes three possible causes for the damage to the system: "*political activities*", "*personal background*", and "*financial background*".

User Assignment – data describes the geographical region of the user and his main tasks.

User Intention – the current classification divides users into "*normal*" and "*dangerous*". It is clear that during one session in the system, the user can change its state. Also, users can have common intentions that do not change over many sessions.

User's Other Intentions – the purpose of the system is not just to identify a malicious user, but also to discover the nature of potential threats. Therefore, it is important to consider other features of user behaviour in the system.

Document – documents have the sources, regions, and the value of the classifier. Also, each requested document has a degree of compliance to the provided request.

Query – users make requests and receive the documents.

The Results of the Simulation

The purpose of the experiments has been to learn to distinguish the user type (*normal* / *malicious*), based on his actions (request of documents) in a certain period of time (committed during the previous sessions in the system). Initially two identical models have been created; one model, trust (ground truth), has been used to simulate user behaviour and intentions, and the second one – inference, designed to determine the presence of anomalies in the current behaviour. Test data (a set of queries of various documents) of 100 sessions inside the system have been generated for each user. By knowing each user type (*normal* / *malicious*) in advance, the behaviour of 192 users has also been simulated.

The analysis of normal behaviour is displayed in Fig. 3. As seen from the figure, for all 100 sessions, the evaluation of the probability of normal behaviour does not fall below 0.9.

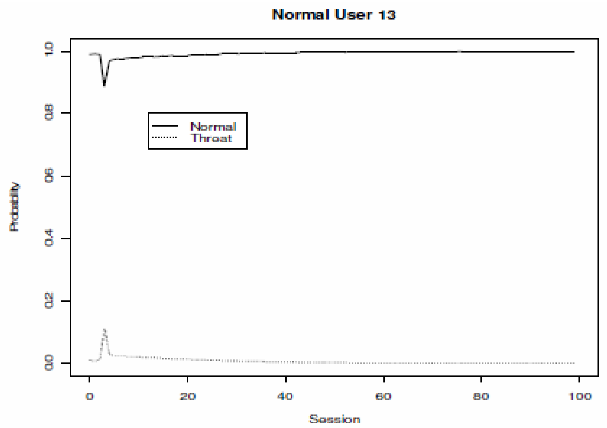

Fig. 3. Normal behaviour of a user

Fig. 4 shows the behaviour of the probability, if the user tries to hide its malicious behaviour.

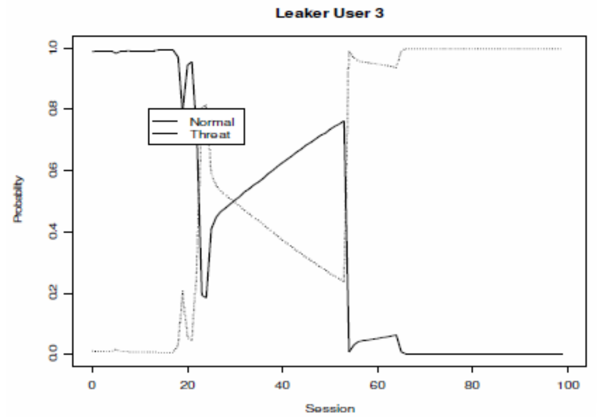

Fig. 4. Anomaly appears after some time of normal behaviour

Fig. 6 presents the case, when a malicious user is deliberately not found. We see that in this case, his behaviour is evaluated as "*normal*" for all the sessions, despite the fact that it is known that he makes unlawful acts.

In contrast to Fig. 4, where the case of delayed detection of anomalous behaviour is displayed, Fig. 6 shows a typical case of a malicious user.

In conclusion, it should be stated that the experiments have been performed on the basis of artificially generated data, and their aim has been to prove the functionality of the overall approach discussed in the paper. In the future, the authors plan to use the data from actual use of the existing system, the ranking of anomalous observed behaviour, an increase in the value used in the description of user characteristics and the security policies close to the real ones.

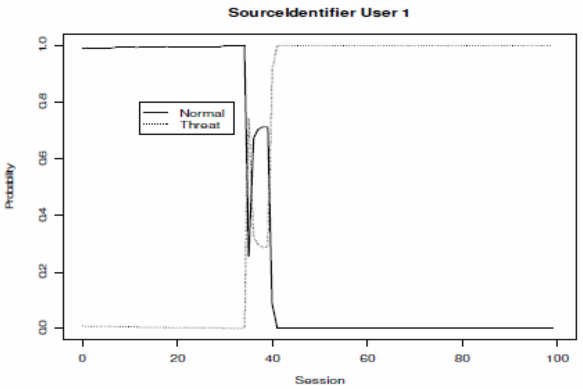

Fig. 5. Typical anomalous behaviour detection

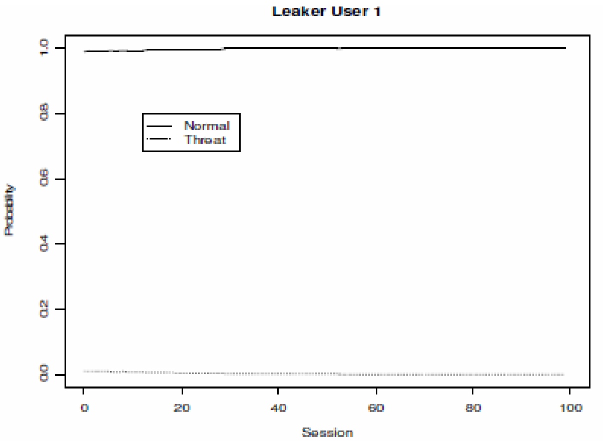

Fig. 6. Undetected anomalous behaviour

As an important area for further research, the authors consider the use of Data Mining methods for the preparation and subsequent analysis of data obtained by Bayesian networks. In particular, it is proposed to create a common ontology document that reflects their internal semantic relations for a more accurate assessment of conformity of the document obtained by the user's request.

C. An Approach Based on the Use of Ontologies

In the work of French scientist L.Razmerita [4], an approach to modelling the user behaviour is described, using techniques from the field of semantic web, as well as ontologies [5].

The approach is based on the use of ontologies, which has recently been gaining popularity because of its flexibility, the quality of the proposed conceptual data management methodologies and knowledge bases, opportunities to disseminate and reuse knowledge.

The described approach considers the modelling of the behaviour of the electron system by using ontologies in the context of knowledge management systems (KMS). The use of ontologies allows describing the features of the target domain in a general way, which will continue to use the resulting description to solve other problems in other applications and other research groups. Additionally, by complicating the ontology it can be approximated to a more accurate description of the behaviour of each individual user.

The very process of creating ontology is a complex and time-consuming task that requires expertise in various fields, such as software creation, object-oriented programming, theory of modelling, artificial intelligence and many others.

In general, the process of creating the ontology [6] consists of three simple steps: collection of knowledge, coding, and integration with existing ontologies. Established as a result of the ontology, it will be structured according to the specification IMS LIP [16]: "The purpose of the specification is to define a set of modules that can be used to import data and retrieve data from a compatible with IMS Information Server".

In order to be able to describe the user, each IMS package of structural information is divided into 10 subgroups. These groups are the following: *Identification, Goal, QCL (Qualifications, Certifications and Licenses), Accessibility, Activity, Competency, Interest, Affiliation, Security Key* and *Relationship*.

Identification describes the personal data of a target person. *Affiliation* includes data on the relationship of a man with the target organization. *QCL* contains a list of skills of a person, his diplomas, certificates and licenses. *Competency* describes other formal and informal human skills, as well as the history of his work. *Activity* describes the human activities associated with training as part of his professional duties. *Accessibility* includes concepts related to possible specific features and user requirements. *Interest* describes a variety of hobbies and interests of a user. The *Goal* concept includes user major and minor goals.

However, is not enough to simply describe the behaviour of these elements; therefore, additional concepts should be introduced. *Behaviour* models the characteristics of the user's interaction with the system. The data for this module comes from the history *log*-files of the user in the system. To construct the heuristics that are suitable for the formation of rules it is necessary to add such things as: *Type_of_Activity, Level_of_Activity* and *Level_of_KnowledgeSharing*. For example, *Type_of_Activity* describes the type of user behaviour in the system – he prefers mostly to read, write, or go unnoticed. Also, each type of behaviour can be classified as *very active*, *active*, *passive or active*. The third type of classification is based on the assessment of the analysis of the spread of the user's knowledge; users are divided into *Unaware, Aware, Interested, Trial,* and *Adopters.*

The next step is to encode the ontology using a formal language, or it can be created using the ontology editor, such as OntoMat, OI-Modeler, KAON, or more often Protégé.

Next, the created ontologies are integrated and used within the overall system.

Implementation of Ontology

At this stage, the mechanisms of modelling and

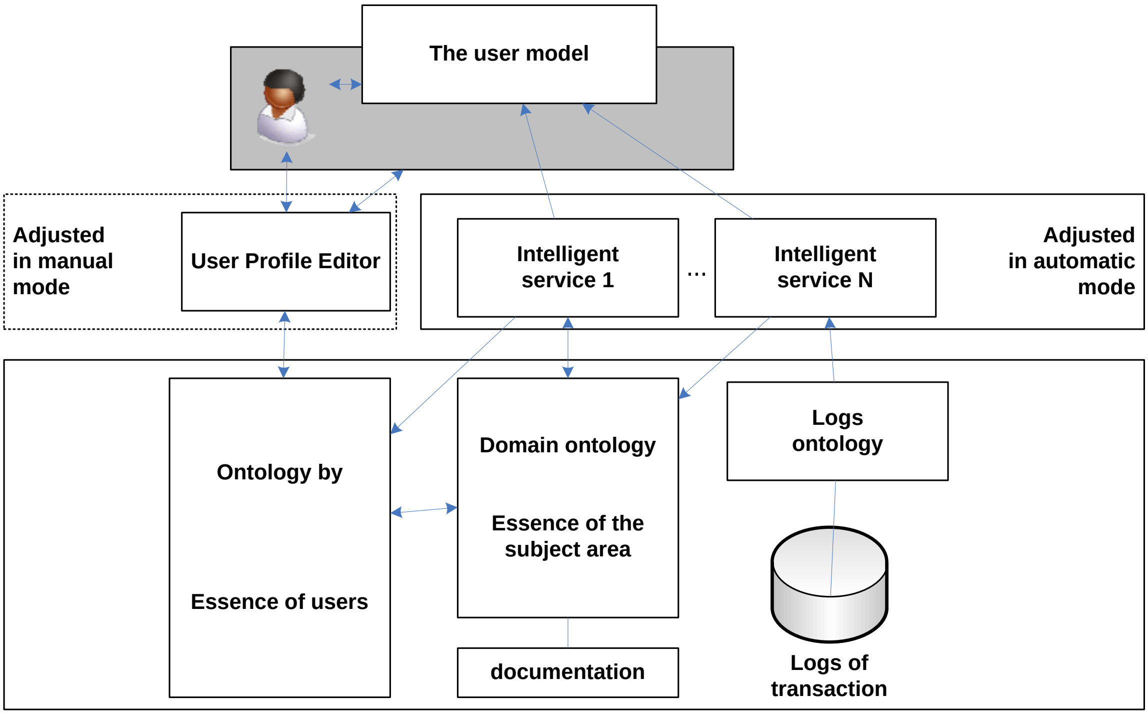

Fig. 7. The implemented architecture

personalization of models are implemented as a set of intelligent services. Fig. 7 presents the architecture of created Ontology-Based User Modelling framework (OntobUMf).

The data used to create a user model have a clear structure and are partly based on the supplied user profiles, as well as, implicitly, on the conclusions of a *category extractor* represented in the form of intelligent service operating in an automatic mode. The task of the *category extractor* is the type of user classification based on the type of its activity in the system. Any action of the system is written to the transaction log.

User Profile Editor – a specialized ontology editor, which is available to end users and controls the description of his personal profile.

Intelligence services – Due to the modular system, OntobUMf can use various types of algorithms as the intelligent services, determining the type (role) of the current user behaviour. The main objectives of the service are the following:

- to update and support the user model based on data from the *category extractor*;
- to provide the individual analysis for specific types of users.

Adaptation and Personalization – the process of adjusting the system to the features of individual user behaviour. For example, it can be adaptive graphical user interface, structure and accessibility of data security policies. The overall objective of this process is to provide the right information at the right time to the right users.

Ontology of the user's interests created as part of the system can be used for various purposes. Depending on the specifics of current task, *category extractor* intelligent service, which classifies users, is implemented.

D. Multi-Agent Approach to the Modelling of the User Behaviour

Another possible approach to creating a behaviour-oriented model for information GRID [8] system is considered in [7]. GRID systems are becoming more common both in research and in the application environment to solve problems of high computational complexity. Current methods of ensuring security in this type of the systems are based on Public Key Infrastructure (PKI) protocols [9]. This approach ensures good conventional protocols, authorization, authentication, delegation and exchange of certificates. However, detection of abnormal activity using these protocols shows the results, which are not good enough. The topic is relevant, and there are several attempts to solve it [10] [11], but at this point sufficiently accurate detection of this type of attack has not been reached.

The analysis of other existing methods for detecting anomalous behaviour of information system user has shown that the methods are ineffective in terms of the specifics of GRID systems. Therefore, the task has been set to develop a new model taking into account the specifics of this type of systems. For example, if a typical user workflow is a consistent implementation of a large number of different teams in the context of the GRID system, the number of tasks is much smaller, but each of the tasks requires significantly more resources.

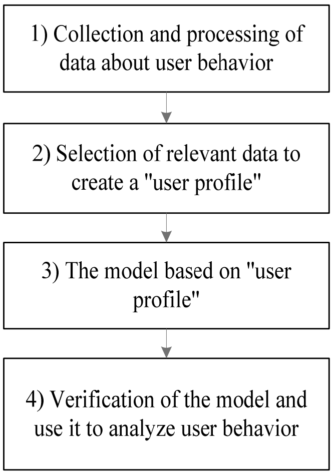
1) Collection and processing of data about user behavior
2) Selection of relevant data to create a "user profile"
3) The model based on "user profile"
4) Verification of the model and use it to analyze user behavior

Fig. 8. Typical steps in the behaviour modelling

Fig. 8 shows the general sequence of actions performed by using a model of user behavior to detect its abnormal activity.

The greatest difference, when using different approaches, is observed in step 3, where they can be used in a variety of approaches. Within the framework of the present research, at this stage it is proposed to use a neural network approach [12], and to implement the entire system using the agent-based approach.

Neural network is a popular approach to classification. In the present research, a multilayer feedforward network has been used. As the input data, the authors of the article have used the following set of characteristics:

{S, ET, CPU, WT, CW, ES, CT, STD, RAM, VM, VO, RB}, where S (*Site*) — the node on which the task is executed; ET (*Execution Target*) — the resource of the node, which performs the task; CPU (*CPU Time*) — CPU time to complete the task using the resource; WT (*Wall Time*) — total task time; CW (*CPUWall = CPU/W*) — the ratio of CPU time to total time of the task; ES (*ExitStatus*) — the status of task completion (success or error); CT (*Creation Time*) — Time of delivery (creation) of the task into the GRID system; STD (*Start Time Difference*) — difference between the start time of a task on the selected resource GRID time systems and sending time to the target GRID system; RAM (*RAM Used*) — the amount of used memory; VM (*Virtual Memory Used*) — the amount of virtual memory used; VO (*Virtual Organization Name*) — ID belonging to a virtual organization; RB (*Resource Broker Hostname*) — ID of the resource task broker.

For each user it is necessary to build his personal neural network, which classifies it as a normal behaviour (1) or abnormal (0).

The system GILDA [14] of the European project EGEE [15] has been used to obtain the data of the users necessary for the experiments. In total, the database consisted of more than 34,000 log entries made by users of the access activities. To monitor all available states of the system in a distributed system, GILDA GridICE has been used, which integrates with the local system resource monitoring.

In the process of software implementation of the test system, the raw data on the behaviour of users have been converted to XML format and then divided into training and test samples in the proportion of 85% to 15%, respectively. The structure of the neural network has been used by a network of direct dissemination of information with one hidden layer, trained by the back-propagation method. To calculate the optimal values of the network structure, A / B testing has been provided, showing that the best configuration is the 20 neurons in the hidden layer, resulting in 85.81% classification accuracy. Also, weights and learning parameters have been assigned the values (η = 0.3, μ = 0.15).

In the final test, the procedure of substituting the user has been used, when, after the initial use of the data on the behaviour of one user, the data classifier has fed the other (illegal) user. The final results have shown that the level of classification of an authorized user is equal to 99.14%, while in the case of substitution by correct classification the level equals 99.30%. The provided results are sufficient for the approach to be used in real systems.

The General Structure of the System, Agent-Based Approach

The most important feature of the GRID systems is their distributed, heterogeneous structure. It is also important to isolate the abnormal activity detection unit from the rest of the system. Third, we need to interact with the system of monitoring to obtain information about the tasks of running users and Certification Centre (CA – Certificate Authority).

All these requirements are satisfied by Agent Based Paradigm implementation. In this case, to detect the presence of abnormality in the actions of a user, it is necessary to use an agent, which encapsulates the personal neural network of a user, and the methods of obtaining the required information from the envelopment analysis.

The monitoring system is the centralized GridICE. To gather information about the tasks performed by users, at each node Grid Resource Information Service (GRIS) runs. In turn, the service domain Grid Index Information Service (GIIS) interacts with local services GRIS, aggregates the information and sends it to a centralized server GridICE (Fig. 9). To implement the model of user behaviour, the following types of agents have been implemented:

- An agent encapsulates a model of GRID user system (User Agent). It is activated after a user-requested action and includes the personalized neural network model of the current user, statistics on his previous behaviour and the presence of anomalous test results in the executed transaction.
- Agent Controller manages the creation and monitors agents, user models. Also, its purpose is to inform the security in the cases, when anomalous activity is detected.

The test system has shown quite good results for its implementation in real projects.

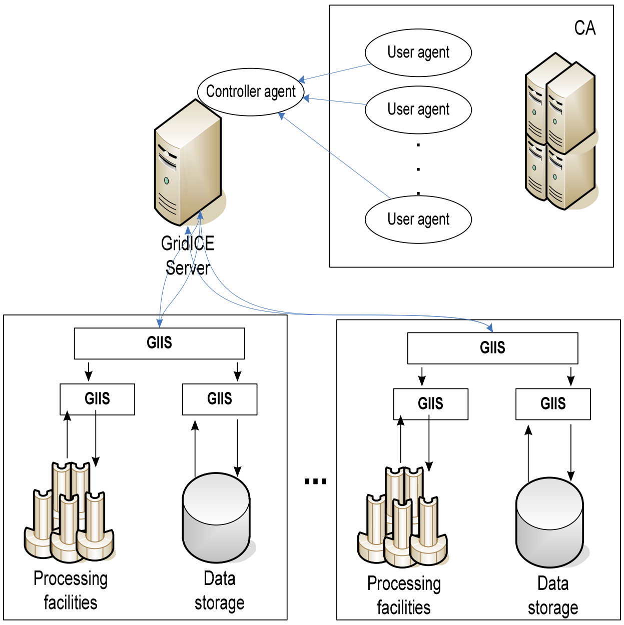

Fig. 9. The interaction of agents with the system

III. Conclusions

The article deals with a variety of methods for modelling user behaviour. This approach is relevant to the intrusion detection system, which monitors system activities for malicious activities of intruders, who use the authorization data of a legitimate user for criminal purposes. Building a behaviour-oriented model, which controls the activities of each user, can prevent the system from this type of invasion, whereas conventional security tools do not provide adequate protection against such attacks.

When you create this type of a system, it is important to evaluate the existing research in this area. To build a Bayesian network, the following models have been used: behavioural models, ontology, neural networks and agent-based approach; each of them has shown good test data results.

References

[1] Infected User Behavioral Modeling for Cyber Defense Exercises: Part I URL http://www.breakingpointsystems.com/resources/blog/cyber-defense-exercises-part1/ Last accessed 2012.05.12

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

[3] The Eighth Annual CSI/FBI 2003 report: Computer Crime and Security Survey.

[4] Razmerita, L., Modeling Behavior of Users in Adaptive and Semantic-enhanced Information Systems: The role of a User Ontology, Authoring of Adaptive and Adaptable Hypermedia workshop, in conjunction with Adaptive Hypermedia Conference, 29 of July-1 August 2008, Hanover 2008.

[5] Bunge. (1977) Treatise on Basic Philosophy: Ontology I - The Furniture of the World. Reidel.

[6] M. Uschold and M. Gruninger, Ontologies: principles, methods, and applications, Knowledge Engineering Review, vol. 11, pp. 93-155, 1996.

[7] A. Shelestov, S. Skakun, N. Kussul, Agent-based approach to implementing a model of user behavior Grid-systems; Space Research Institute NASU-NSAU; Інформатика, кібернетика та обчислювальна

техніка, вип. 9 (132), Донецк, ДонНТУ, 2008. – С.8-14. ISSN: 1996-1588;

[8] Foster Ian The Grid: Blueprint for a New Computing Infrastructure. — Morgan Kaufmann Publishers. — ISBN 1-55860-475-8 , 677p, 1999

[9] Adams C., Lloyd S. Understanding PKI: Concepts, Standards, and Deployment Considerations. 2nd ed. Addison-Wesley, 2000.

[10] Seung-Hyun K., Kyong H.K., Jong K., Sung-Je H., Sangwan K. Workflow-Based Authorization Service in the Grid. J. of Grid Computing, 2004, Num. 2, P. 43–55.

[11] 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), May, 2003.

[12] Haykin S. Neural Networks: a comprehensive foundation. Upper Saddle River, New Jersey: Prentice Hall, 1999.

[13] P. A. Osipov and A. N. Borisov; Abnormal action detection based on Markov models; Automatic Control and Computer Sciences; Volume 41 / 2007 - Volume 45 / 2011; ISSN 0146-4116 (Print) 1558-108X (Online); May 05, 2011.

[14] https://gilda.ct.infn.it/ Last accessed 2012-04-24

[15] http://www.eu-egee.org/ Last accessed 2012-04-24 IMS Learner Information Packaging Information Model Specification, Final Specification; Version 1.0; http://www.imsglobal.org/profiles/lipinfo01.html Last accessed 2012-04-24

Arkady Borisov holds a degree of Doctor of Technical Sciences in Control of Technical Systems and a habilitation degree in Computer Science.

He is a Professor of Computer Science at the Faculty of Computer Science and Information Technology, Riga Technical University (Latvia). His research interests include fuzzy sets, fuzzy logic and computational intelligence. He has 205 publications in the fields of computer science and information technology.

He has supervised a number of national research grants and participated in the European research project ECLIPS.

Pavel Osipov, Mg.sc.comp., Doctoral Student of the Institute of Information Technology, Riga Technical University. He received his Master Degree Diploma in Transport and Telecommunication Institute, Riga.

His research interests include web data mining, machine learning and knowledge extraction. Research activities are mainly focused on different aspects of user behaviour modelling. A new area of interests is related to the exploration of application of Python programming language to all steps of scientific research.

Pāvels Osipovs, Arkādijs Borisovs. Pieejas informācijas sistēmu lietotāju uzvedības modeļu radīšanai

Rakstā aprakstītas trīs dažādu elektroniskās informācijas sistēmas lietotāju uzvedības modeļu veidošanas pieejas. Atkarībā no modeļa lietošanas mērķa un apskatāmo sistēmu struktūras īpašībām profila izveidošanā var izmantot dažādus algoritmu veidus. Pētījumā apskatītas trīs pieejas: neironu tīklu un aģentu apvienošanas pieeja, Baijesa tīkli un ontoloģijas. Katrai ir savas priekšrocības un trūkumi, funkcijas un iespējamais pielietojums sarežģītu elektronisko sistēmu mūsdienu infrastruktūrā. Vislabākos rezultātus katra pieeja uzrāda tās informācijas sistēmas ietvaros, kurai tas tika izstrādāts. Izmantojot Baijesa tīklus, var modelēt sarežģītas motīvu un visu procesa dalībnieku iespēju mijiedarbību. Tāpat tie ļauj izmantot gan automātisku datu izdalīšanu no lietotāju aktivitātes logiem, gan eksperta metodes izmantošanu modeļa mezglu svaru un sadalījumu, kas apraksta dažādus dalībnieku raksturlielumus, parametru uzstādīšanai. Ar ontoloģiju iespējams ļoti detalizēti aprakstīt dažādus lietotāja uzvedības aspektus. Pētījuma ietvaros netika izskatīts drošības nodrošināšanas mērķis, lietotāja uzvedības modelis tika sastādīts intelektuālai rekomendējošai sistēmai. Ontoloģijas uzrādīja labu rezultātu, ļaujot aprakstīt un ņemt vērā visas svarīgākās katra lietotāja interešu īpatnības. Trešā pieeja ir neironu tīkli, kas kombinēti ar multiaģentiem. Mērķa sistēmai šajā gadījumā ir plaša struktūra, kas prasīja aģentu izmantošanu, lai sniegtu informāciju neironu tīklam. Pats neironu tīkls tika izmantots kā vienkāršs klasifikators, kas atbild uz jautājumu, vai pašreizējās lietotāja darbībās ir atrodami ļaunprātīgi vai anomāli sistēmas izmantošanas mērķi. Šis pētījums ir daļa no projekta, kas paredzēts, lai nodrošinātu drošību necentralizētai informācijas sistēmai pret uzbrukumiem, kurus raksturo "anomāla uzvedība", kad esošo lietotāja kontu ļaunprātīgi izmanto cita persona. Tādā gadījumā tipiskas lietotāja uzvedības profila analīze, izmantojot informāciju par iepriekšējiem darba seansiem ar sistēmu, ļauj noteikt anomālijas un signalizēt drošības sistēmai par ievērojamu atšķirību no ierastā darbības veida.

Павел Осипов, Аркадий Борисов. Подходы к созданию шаблонов поведения пользователей информационных систем

В статье рассмотрены три различных подхода к построению модели поведения пользователя электронной информационной системы. В зависимости от цели использования такой модели и от особенностей структуры целевой системы, для построения профиля возможно использование большого количества разнообразных алгоритмов. Рассмотрены следующие подходы: алгоритмы, использующие нейронные сети и агентный подход, а также - Байесовские сети и онтологии. Каждый из подходов имеет как преимущества, так и недостатки, а также - особенности применения в условиях современной инфраструктуры сложных электронных систем. Наилучшие результаты каждый из подходов показывает в рамках той целевой информационной системы, для которой он был создан. С помощью Байесовских моделей можно моделировать сложное взаимодействие мотивов и возможностей всех участников процесса функционирования системы. Они позволяют использовать как автоматическое выделение данных из логов активности пользователей, так и использование экспертного метода для настройки весов узлов модели и параметров распределений, описывающих те или иные характеристики участников. В свою очередь, онтология позволяет детально описать различные аспекты поведения пользователя. Рассмотренное исследование имело цель, отличную от обеспечения безопасности; модель поведения пользователя в этом случае строилась для интеллектуальной рекомендательной системы. Онтологии показали хороший результат, позволив описать и учесть все важные особенности интересов каждого конкретного пользователя. Третий подход предусматривает использование нейронных сетей совместно с парадигмой многоагентных систем. Целевая система в этом случае имела распределённую структуру, что обусловило использование агентов для поставок информации нейронной сети. Сама нейронная сеть использовалась в качестве обычного классификатора, отвечающего на вопрос: есть ли в текущих действиях пользователя наличие злонамеренности или аномальность. Данное исследование произведено в рамках проекта по обеспечению безопасности распределённой информационной системы от атак типа «Аномальное поведение», когда под учётной записью легитимного пользователя систему злонамеренно использует другой человек. В таком случае анализ профиля типичного поведения пользователя в рамках предыдущих сеансов работы с системой позволяет обнаружить аномалии и сигнализировать системе безопасности о заметном отличии в шаблоне поведения.