VERSITA
Information Technology and Management Science
doi: 10.2478/v10313-012-0020-9
2012 /15

The Evaluation of Capabilities of Intellectual Intrusion Detection Systems for the Use in Web-Based Information Systems

Pjotrs Dorogovs[1], Andrejs Romanovs[2], [1-2]*Riga Technical University*

***Abstract* – Nowadays with vast growing number of network information systems and their integration not only into work but also into people's private life, security assurance of industrial and private information assets is becoming an extremely sensitive and topical issue. There is a huge number of available non-commercial (free of charge) and commercial methods for information protection from unauthorized access of undesirable individuals. Currently, studies in the field of information security focus on the use of various intellectual data mining techniques for building an intellectual information security system. Such security systems roughly (for the purpose of this paper) can be divided into intrusion protection and intrusion detection systems – IPS and IDS, respectively [1].**

***Keywords* – information security, intrusion detection, Web-based system security, system modelling**

I. Introduction

Recently, the problem of ensuring network security has been the stopping factor for development of the Internet applications. Being one of the most important network technologies, intrusion detection systems still have some drawbacks. By utilizing intellectual data mining technologies for Web-based applications, some problems of intrusion detection can be solved. This paper will describe common architecture of intrusion detection systems for Web applications; moreover, main modules of such a system, data flows and functional issues will be pointed out. Besides, some principles of common intellectual data mining technologies that can be used for Web IDS will be described.

II. Overview

In general, intrusion protection systems include any available method or recommendation that prevents attackers from gaining access to a secured network, system or information asset. Most common ways to ensure high availability intrusion protection system is the use of any kind of firewall or anti-virus software. Each type of IPS has different levels of provided protection. Sometimes it is even advisable to build an IPS containing more than one security solution.

The intrusion detection system, in turn, may be considered a type of security assurance methods both for information systems and computers. Such a system should make the comprehensive analysis of the gathered information of computer, network or information system activities to proactively identify potential security breaches that may include both attacks from inside and outside of the protected perimeter. The fact that data and systems cannot always be protected from outside intruders in the contemporary Internet environment using ordinary security mechanisms, such as passwords and file security, leads to a range of issues.

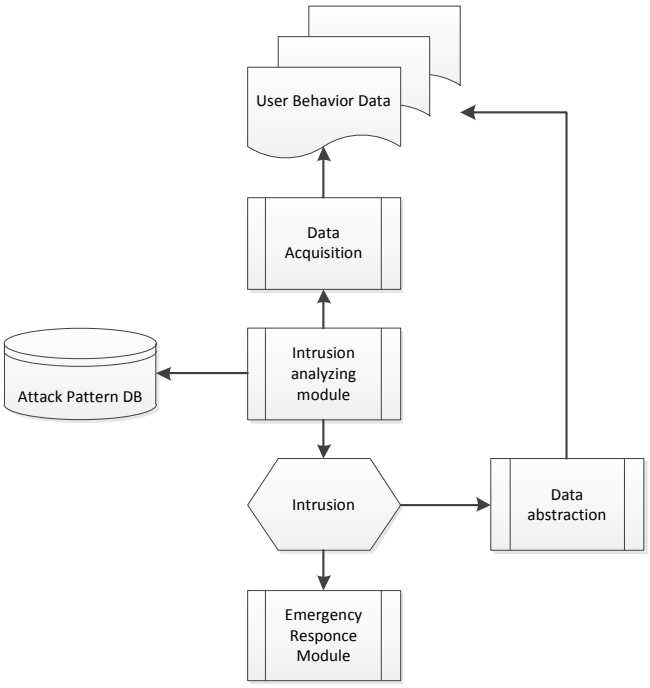

Fig. 1. General module of the intrusion detection system

Intrusion detection is a method to effectively protect the host system from possible attacks. As a supplement to firewalls, intrusion detection systems can provide additional options for systems dealing with network attack, expand the ability of a system administrator to assure information security, and promote the completeness of information security in three main information security aspects.

Mainly IDS consists of data acquisition, intrusion analyzing and emergency response modules. The structure is shown in Fig. 1.

The acquisition module collects the log files of the analyzed system and Web, network traffic, etc.

Intrusion analyzing module is responsible for the analysis of collected log files using intellectual data mining techniques, such as Hidden Markov Models, Audit Trail Pattern Analysis, K-Nearest Neighbour Classification, Principle Component Analysis, Association Rules, etc., and integrity analysis and later responds whether it is intrusion.

Emergency response module has to initialize security actions depending on the provided information regarding possible intrusion. Such security actions may include software based activities, such as stopping of attacked processes, initializing back-up procedures etc., as well as hardware activities – collapse of the network connection, closing of Web services, etc.

Currently, all intrusion detection systems available on the market fall into two categories – network-based systems, which are placed in the network nearby the system that is being monitored and that examine network traffic, and host-based systems, which actually run in the system being monitored and that examine the activity of the monitored system. Most recent type of intrusion detection systems resides in the operating system kernel and monitors the activity at the lowest available level of the protected system.

Network- and host-based IDSs bring very similar advantages (Table). Both of them are very well suited for outsider deterrence. Network-based systems are able to warn attackers regarding their illegal actions, thus working as a buffer for inexperienced hackers showing them that they are not as safe as it seems. On the contrarily, host-based systems work on the assumption that people that are aware of constant monitoring of their actions are less likely to commit misuse. Although both types of systems are able to detect a vast variety of intrusion actions, the former is more oriented exactly to network activities while the latter is able to detect more insider actions. Furthermore, both systems can react and even alert security personnel of possible misuse.

TABLE I

HOST- AND NETWORK-BASED SYSTEMS

Benefit	**Host-based systems**	**Network-based systems**
Deterrence	Strong deterrence of insiders.	Strong deterrence of outsiders.
Detection	Strong insider detection. Weak outsider detection.	Strong outsider detection. Weak insider detection.
Response	Weak real-time response. Good for long-term attacks.	Strong response against outsider attacks.
Damage Assessment	Excellent for determining the extent of compromise.	Very weak damage assessment capabilities.
Attack Anticipation	Good at trending and detecting suspicious behaviour patterns.	None.
Prosecution Support	Strong prosecution support capabilities.	Very weak because there is no data source integrity.

In recent years, a vast majority of research activities in the area of anomaly detection have been focused on studying the behaviour of programs and the creation of their profiles based on system call log files. Until now, a simple anomaly detection method based on monitoring system calls initiated by the active and privileged processes is widely used.

The profiling of the behaviour of the end user is not less important aspect of data protection than the profiling of the software activities. This method is effective in detecting internal attacks that constitute one-third of the corporate system security. On the other hand, it is difficult to take into account the difference between the behaviour of end users and to build profiles of their activities in comparison with the building of a profile of program behaviour. Hackers can even try to adapt their behaviour to fool IDS systems [4].

III. DATA MINING TECHNIQUES

The profiling of the software and network activities is not less important aspect of data protection than the profiling of behaviour of the end user. This method is effective in detecting external attacks that constitute almost two-thirds of the corporate system security.

A. Hidden Markov Model (HMM)

A particularly powerful method that uses a fixed number of states is a hidden Markov model, which is widely used in speech recognition, as well as in the simulation of DNA sequences [6], [7]. The hidden Markov model (HMM) describes a double stochastic process. HMM states represent some unobservable conditions of the system being modelled. In each state there is a certain probability of creating one of the possible system outputs and a separate probability indicating a possible next state.

The time to check each network call depends on the size of the model and size of the current list of valid states. This list tends to be very small, if it consists of only "normal" tracks of network activities and contains all the possible states after the discovery of anomalies.

B. Audit Trail Pattern Analysis

Algorithm to build a table of fixed-length patterns of normal network activities is very simple. From the sequence of data passed through the analysis module, all the unique subsequences are retrieved (models of a given length k). This is achieved by using a sliding window of length k for all the input data, followed by writing all occurring subsequences.

Pattern-matching technique is similar to the pattern-generation technique. The k-length window is moved through the sequence, recorded during the actual operation of the system with the network flow. Each packet is checked for a *match*, i.e., whether there is a pattern that corresponds to the subsequence in the window.

To detect an attack, at least one of the subsequences generated by the attack itself should be classified as an anomaly. However, as shown by experimental results, full compliance cannot always be achieved [5]. Thus, the threshold is defined so that only the sequences with a number of occurrences above this threshold are considered suspicious.

C. K-Nearest Neighbours

Instead of analyzing the local ordering of network packets, data on the frequency of those packets can be used to characterize network behaviour. Using the metaphor of text processing, each network packet is treated as a "word" in a long document, and a set of network packets generated by each network connection is seen as a "document". This

analogy enables us to use the full range of well-developed methods of text processing [8] for the problem of intrusion detection. One such method is the method of K-nearest neighbour classification [9].

By analogy with text categorization, each network connection initially appears as a vector where each entry is a network packet during the connection. Ranking techniques, such as the frequency weighting and $tf \cdot idf$, are used to determine the values of vector elements. To classify a new connection as normal or intrusive, kNN classifier computes the similarity between the new connection and each instance of training data, and uses the class labels of *K* closest neighbours to define the class of the new connection. This approach contributes to the underlying assumption that the connections belonging to one class will be collected in a single cluster in the vector space.

D. Association Rules

The above-mentioned data mining methods are widely studied and discussed in many articles and scientific experiments. In turn, a method for association rule construction being one of the most popular approaches among various data mining methods is relatively rarely used in the intrusion detection sphere.

Initially, association rule induction method has been proposed to be used for the so-called problem of market basket analysis, which aims to find patterns in behaviour of supermarket consumers. In particular, Boolean associative rules aim to identify sets of products (items) that are often bought together. Discovered rules may, in particular, tell us that people who buy butter and milk will also buy bread.

In other words, the purpose of the analysis is to identify the following dependencies: if a certain set of elements of X appears in the transaction, then on the basis of this we can conclude that a different set of Y will also appear in this transaction. Establishing such relationships enables us to find very simple and intuitive rules.

The association rule induction algorithm can be used for two-phase (learning and detection phases) intrusion detection system. Taking into account that during the learning phase the system learns and analyses raw network data provided and then, on the basis of these rules, creates a network security model that will subsequently be used for a real intrusion detection, it can be concluded that the construction of proper rules of normal behaviour of the system is a key factor for a network intrusion detection system.

IV. WEB-BASED SYSTEM SECURITY

For a web-based system, the displacement of intrusion detection system (Fig. 2) is one of the vital issues to be solved to ensure security of information assets. Comparing to end-user applications, where IDS is likely to be deployed on the host level, which is otherwise defenceless (i.e., Windows-based machines running previous versions of OS are unable to create even simple logs that later can be processed by an off-line IDS), WEB-based systems should be protected on the network level rather than on the host level. In this case, the intrusion detection system will be more effective when placed in the network perimeter, i.e., just behind and/or before the firewall, on links to partners, etc. Otherwise, it can be placed on the corporate WAN backbone, where it is possible to monitor all the traffic that attempts to enter a corporate network. In special cases to ensure high-end protection of valuable information storage or processing units solution of isolation of critical infrastructure into a different network segment with its own IDS is considered to be state-of-art technology in the field of information security [3].

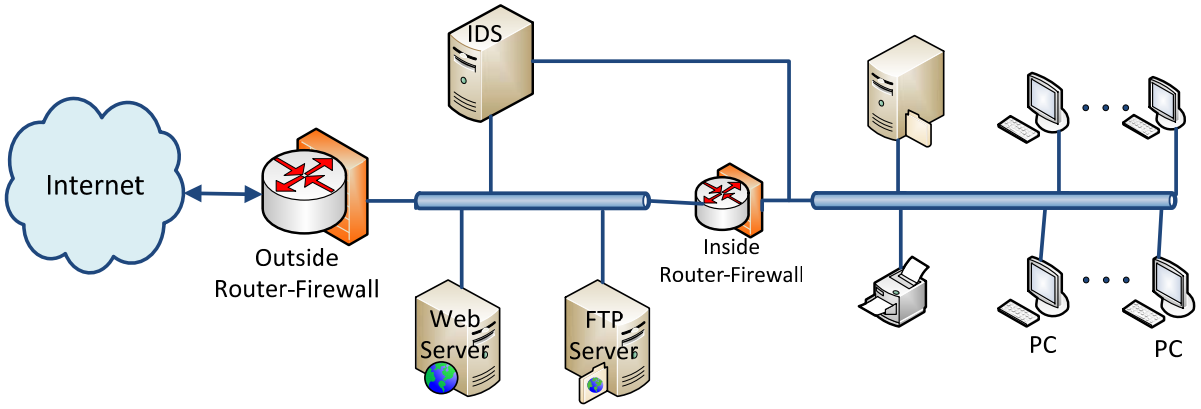

Fig. 2. Displacement of the intrusion detection system

Fundamentals for implementation strategies for intrusion detection systems to build a trustable defence system were put forward in 2000. Among many other valuable issues it should be noted that implementation of firewalls between areas of the network with different requirements (i.e. between internet-intranet, between users-servers etc.), usage of network vulnerability scanners to double check firewalls and to find security holes that intruders can exploit, usage of host policy scanners to make sure they conform to accepted security practices and finally usage of symbiosis of NIDS, other packet sniffing utilities and host-based virus scanners to flag successful intrusions may significantly improve the overall level of information security of web-based system [2].

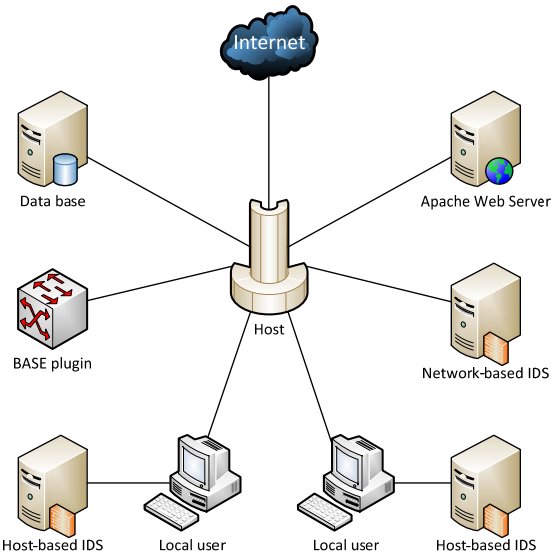

Fig. 3. Snort IDS example

Most of modern intrusion detection methods are directed to the proactive analysis of the already conducted attacks and to the creation of new rule sets for detection of next attack of the same type on the basis of the previously generated rules. Scope and efficiency in such a case will be limited with those rules for specific types of attacks. Nevertheless, enormous traffic caused by a new attack cannot be detected. Consequently, it is crucial to perform the fast analysis of anomalous traffic instead of a detailed one to make it possible to determine possibility of incoming anomalous traffic. Usually the network traffic analysis consists of the following basic functions: primitive network traffic data, integration of traffic data and detection of anomalous network behaviour. The main concern with the network traffic analysis is not just the traffic count but the definition, which network should analyse traffic features so that the actual collecting method and types are determined. Method for detection of abnormal network behaviour is analogous to the intrusion detection system. It detects and analyses network traffic that consists of attacks based on network traffic patterns of well-known attacks. Another method for traffic classification is based on the modelling of normal behaviour of network activities. Both methods require the modelling of network traffic and the analysis of related functionality for abnormal traffic. Generally well-known tools, such as Ntoop, are used for traffic analysis. Besides, TcpDump, IPmon and Snort tools can be used.

Snort for example is a behaviour-based and rule-based NIDS that demonstrates outstanding performance in real-time traffic analysis and packet log analysis (Fig. 3). It can be utilized for the protocol analysis, examination of packet content, pattern matching, port scan, CGI attacks, buffer overflows, etc. It uses a very flexible rule language consisting of a module plug-in structure to catch traffic. Three main Snort functions are the following:

- It can be used as a packet sniffer such as TcpDump;
- Network traffic debugging is available based on the internal packet logging function;
- It shows good NIDS functionality. Snort is a packet-sniffing tool that uses the packet capture library of Libpcap.

Snort recognizes sniffed packets and compares them with pre-defined detection rules via a pre-processor and a detection engine to detect the intrusion. Rules for Snort can be easily created by users that later can be applied as plugin operations with different alert logs and pre-processors. Nevertheless, due to simple pattern-matching the possibility of false-positives is quite high and the detection of new attack types is almost impossible.

As mentioned previously, usually the IDS is deployed near the web server and it monitors the network activities by performing the protocol analysis and pattern matching. In other words, IDS should reconstruct HTTP headers and payload from captured packets, and identify attacks by comparing traffic to signatures of attacks or behaviour profiles.

Such mechanisms as SSL (Secure Socket Layer) (Fig. 4) or its successor TLS (Transport Layer Security Protocol) were introduced as solution for secure communication and data transfer over the Internet. These protocols, first of all, allow authentication of servers and users and, secondly, considerably contribute to safeguarding confidentiality, integrity and availability of data. For encrypted traffic, simple intrusion detection systems need a deposited private key. Alternatively, they will have to monitor traffic just after decryption.

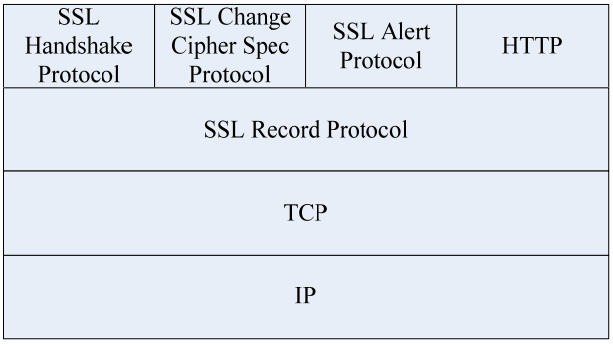

Fig. 4. SSL protocol stack

These conventional approaches are problematic from the perspective of key management and network configuration and tuning. However such approaches are becoming more and more popular, taking into account the rising popularity of web systems and applications that require secured communication between an end-user and a server. Thereby, web-application server administrators are faced with the dilemma either to provide secured services using SSL/TLS protocols but with a less secure system itself because of lack of IDS monitoring or vice-versa.

V. Conclusions

Intrusion detection system is an important addition to commonly used firewalls, encryption of network traffic, access control solutions and other traditional information security approaches. The paper has described only a few possible intellectual data mining algorithms that can be integrated into a contemporary intrusion detection system. It should be mentioned that the use of such algorithms allows information security managers to step beyond traditional opinions regarding web security since compared to other

traditional intrusion detection systems, intellectual systems are safer and more efficient.

REFERENCES

[1] P.Dorogovs, A.Borisovs, A.Romanovs. *Building an Intrusion Detection System for IT Security Based on Data Mining Techniques*, Scientific Journal of Riga Technical University, Series 5, Computer Science, Vol.40, Information Technology and Management Science, pp.43-48, 2011.

[2] Joseph S. Sherif, Tommy G. Dearmond, *Intrusion Detection: Systems and Models,* Proc.of the 11th IEEE Intern.Workshop on Enabling Technologies for Collaborative Enterprises, WETICE'02, pp.115-136, 2002.

[3] W.Lee, Salvatore J.Stolfo, *A Framework for Constructing Features and Models for Intrusion Detection Systems*, ACM Transactions on Information and System Security (TISSEC), Vol.3, Issue 4, pp.227-261, 2000.

[4] Charu C. Aggarwal, Z.Sun, Philip S. Yu, *Fast Algorithms for Online Generation of Profile Association Rules*, IEEE Transactions On Knowledge And Data Engineering, Vol. 14, No. 5, pp.1017-1028, 2002.

[5] Steven A. Hofmeyr, Stephanie Forrest, and Anil Somayaji. *Intrusion detection using sequences of system calls.* Journal of Computer Security, 6(3):151–180, 1998.

[6] L. R. Rabiner. *A tutorial on Hidden Markov Models and selected applications in speech recognition*. Proceedings of the IEEE, 77(2), pp.257–286, 1989.

[7] L. R. Rabiner and B. H. Juang. *An Introduction to Hidden Markov Models*. IEEE ASSP Magazine, pp. 4–16, January 1986.

[8] K. Aas and L. Eikvil, *Text Categorisation: A Survey*, http://citeseer.nj.nec.com/aas99text.html, 1999.

[9] Y.H. Liao and V. R. Vemuri, Use of K-nearest Neighbor Classifier for Intrusion Detection. Computer & Security, vol. 21, No 5, pp. 439-448, 2002.

Pjotrs Dorogovs is a doctoral student at the Department of Modelling and Simulation, Riga Technical University (Latvia). He received his Bachelor Degree in Information Technology from Riga Technical University in 2005. He obtained his Master Degree in IT Project Management (MSc.ing.) from Riga Technical University in 2008. His research interests include IT security and IT governance. Currently, he is an Acting Deputy Chief of the Information Centre of the Ministry of the Interior of the Republic of Latvia. Since 2006 he has been participating in monthly large-scale IT system management forum taking place mostly in Brussels organized by the European Parliament. He has managed the implementation of some large-scale IT systems for Latvian law-enforcement authorities, including putting into production the Schengen information system. He is a member of the IEEE. He has participated in several international scientific conferences and research projects with scientific publications in the field of ICT. E-mail: pjotrs.dorogovs@rtu.lv

Andrejs Romanovs, Dr.sc.ing, Associate Professor and Senior Researcher at the Information Technology Institute, Riga Technical University. He has 25 years of professional experience in teaching postgraduate courses at RTU and developing more than 50 industrial information systems as an IT project manager.

His professional interests include modelling and design of management information systems, information systems for healthcare, IT security and risk management, IT governance, integrated information technologies in business, as well as education in these areas.

A.Romanovs is a senior member of the IEEE and LSS, Council Member of RTU ITI. He is the author of 2 textbooks and more than 30 papers in scientific journals and conference proceedings in the field of Information Technology. He also participated in 25 international scientific conferences, as well as in 7 national and European-level scientific technical projects. E-mail: andrejs.romanovs@rtu.lv.

Pjotrs Dorogovs, Andrejs Romanovs. Intelektuālo sistēmu pielietošanas izpēte ielaušanās atklāšanas izmantošanai WEB-bāzētās informācijas sistēmās

Darba mērķis ir izpētīt intelektuālo ielaušanās atklāšanas sistēmu iespējamo pielietojumu WEB-bāzētām informācijas sistēmām. Tikai pirms dažiem gadiem tīkla drošības nodrošināšanai aktuāla problēma bija apstāšanās faktors interneta lietojumprogrammatūras attīstībai. Tā ir viena no svarīgākajām tīkla tehnoloģijām, bet Ielaušanās atklāšanas sistēmām joprojām ir nopietni trūkumi. Dažus no tiem var atrisināt, izmantojot intelektuālās datu ieguves tehnoloģijas tīmekļa lietojumprogrammās. Darbā ir izpētīti un aprakstīti kopējie arhitektūras risinājumi ielaušanās atklāšanas sistēmās, kā arī ir atspoguļoti to galvenie moduļi, datu plūsmas un funkcionēšanas principi. Papildus tiek apskatītas visizplatītākās intelektuālās datu ieguves tehnoloģijas, kurus ir iespējams izmantot ielaušanās sistēmās, tie ir: paslēptie Markova modeļi (Hidden Markov Model (HMM)), fiksētā garuma Auditācijas pierakstu analīze, K-tuvāko kaimiņu metode, kā arī asociatīvo noteikumu metode. Nepārprotami, ielaušanās atklāšanas sistēmas ir būtisks papildinājums plaši izmantotajiem ugunsmūriem tīkla trafika šifrēšanai, piekļuves kontroles risinājumiem un citām tradicionālajām informācijas drošības pieejām. Darbā tika apskatīti tikai daži iespējamie intelektuālie datu ieguves algoritmi, kurus var integrēt Ielaušanās atklāšanas sistēmās. Ir jāatzīmē, ka šādu algoritmu izmantošana ļauj informācijas drošības pārvaldniekiem spert soli ārpus tradicionālajiem priekšstatiem par interneta drošību, jo, salīdzinot ar citām tradicionālajām ielaušanās atklāšanas sistēmām, intelektuālās sistēmas ir drošākas un efektīvākas.

Петр Дорогов, Андрей Романов. Исследование возможностей использования интеллектуальных систем обнаружения вторжений в ВЕБ-ориентированных информационных системах

Целью работы является исследование возможного применения систем интеллектуального обнаружения вторжений в ВЕБ-ориентированных информационных системах. Несколько лет назад актуальность проблемы обеспечения сетевой безопасности была ключевым фактором, ограничивающим дальнейшее развитие прикладных программ в интернете. На сегодняшний день, являясь одной из важнейших сетевых технологий, системы обнаружения вторжений имеют значительные недостатки. Некоторые из них могут быть нивелированы, используя методы интеллектуального анализа данных в сетевых приложениях. В работе рассматриваются общие технологические решения в системах обнаружения вторжений, а также описываются главные модули, потоки данных и принципы функционирования. Дополнительно рассматриваются такие распространённые технологии интеллектуального анализа данных, как скрытые Марковские модели, анализ аудитных данных фиксированной длины, алгоритм, основанный на анализе К-ближайших соседей, а также построение ассоциативных правил. Несомненно, системы обнаружения вторжений являются значительным дополнением к широко используемым решениям – брандмауэрам, шифрованию сетевого трафика, решениям разграничения доступа и другим традиционным подходам по обеспечению информационной безопасности. В работе рассматриваются только некоторые возможные алгоритмы интеллектуального анализа данных, которые могут быть интегрированы в системы обнаружения вторжений. Необходимо отметить, что использование таких алгоритмов даёт возможность управляющему информационной безопасностью расширить возможности традиционных подходов к безопасности в Интернете, так как в сравнении с другими системами обнаружения вторжений, интеллектуальные системы - безопасней и эффективней.