Кропачьов А. В., Зуєв Д. О.

UDC 681.3.093:044.3

BASIC APPROACHES OF DEVELOPMENT OF DATA CENTER PROTECTION SYSTEMS

A. V. KROPACHEV

Bell Integrator USA Automation Solution Department Manager USA, Colorado

D. O. ZUEV

Independent Consultant Lead Arcitect, Network and Cloud USA, Colorado

E-mail: root@dzuev.pro, artem.kropachev@gmail.com

Abstract. *Data Center cyber-protection methods based on host-based intrusion prevention systems and network based intrusion prevention systems were considered. Basic algorithm of intrusion prevention system functioning and operational readiness evaluation which includes objects of analysis, procedures and evaluation indicators was discussed. It was shown that procedures to be done by Data Center cyber-protection system are identification of the event, signatures database management and denial management. Evaluation of intrusion prevention system efficiency was proved to be based on errors' numbers and scalability. Thereby it should include accuracy, robustness, performance and scalability parameters. Main prevention systems which show model of detection systems interaction with monitored environment events were discussed. Specifically detection strategy based classification which includes cyber-attack signatures analysis, anomalies analysis, hybrid strategy, detection system behavior based classification which includes active behavior, passive behavior, monitored environment based classification which includes local network, global network, hybrid environment, detection system architecture based classification which includes centralized architecture, distributed architecture, hierarchical architecture, detection system performance based classification which includes real time analysis, offline analysis were analyzed. It was mentioned that anomaly-based systems development has to be supervised by operators and adapted to the parameters of the Data Center network. They were divided to three groups: statistical modeling, knowledge based modeling and modeling based on machine learning techniques. It was mentioned that cyber-threats could be modeled as process of transmission of data in hidden channel that change state of some functional node of Data Center. Unified mathematical model of intrusion detection system work which includes states of the infrastructure functional nodes, events involved in a system and transition between the states caused by those events was proposed.*

Keywords: *Data Center, intrusion prevention system, robustness, hybrid environment, anomaly-based system, machine learning, mathematical model.*

Introduction. Strategy for Data Center protection is based on system of perimeter security that incorporates different intrusion prevention systems (IPS): Data Center's security policies as a basis for firewalls and access lists

(ACLs), host-based intrusion detection system (HIDS) and network-based intrusion detection system (NIDS). Development of NIDS which deals with deal with large class of external attacks is crucial and most complicated stage of perimeter security implementation

Basic algorithm of IPS functioning and operational readiness evaluation is shown at Fig. 1. It consists from objects of analysis, procedures and evaluation indicators. IPS works with a global network's data which should be sorted out for legitimate data, attempts of unauthorized access, malware and cyber-attacks (CA) signatures. Thereby, procedures to be done are:

- identification of the event;
- signatures database management;
- denial management.

Identification procedure recognize event and send to the database CA signatures, malicious applications code samples, system vulnerabilities and critical elements of its topology. Denial management procedure analyzes correlation results and forms alerts or allows execution of the program code. Mathematical model of illegal event identification procedure implementation implies receiving of further results set: true positives (TP) intrusion attempts, true negatives (TN) which corresponds to legitimate code, false positives (FP) for legitimate events incorrectly classified as attacks, false negatives (FN) for intrusion event that is not recognized.

Evaluation of IPS efficiency is based on TP, TN, FP, FN quantities, flexibility of the system and hardware-software complex resources. Thereby following prevention systems' features should be considered [1-5]:

- accuracy;
- robustness;
- performance;
- scalability.

IPS accuracy parameter is based on quantification of TP/TN and FP/FN ratios; robustness measures fault tolerance (FP value) to evaluate impact of the most common mistakes and develop fault tolerant IPS; performance determines ability to process data in real time which depends on the detection strategy; while scalability shows IPS ability for scaling adaption to new monitoring platform which is very important for modern Data Centers' infrastructures which tend to evolve more rapidly.

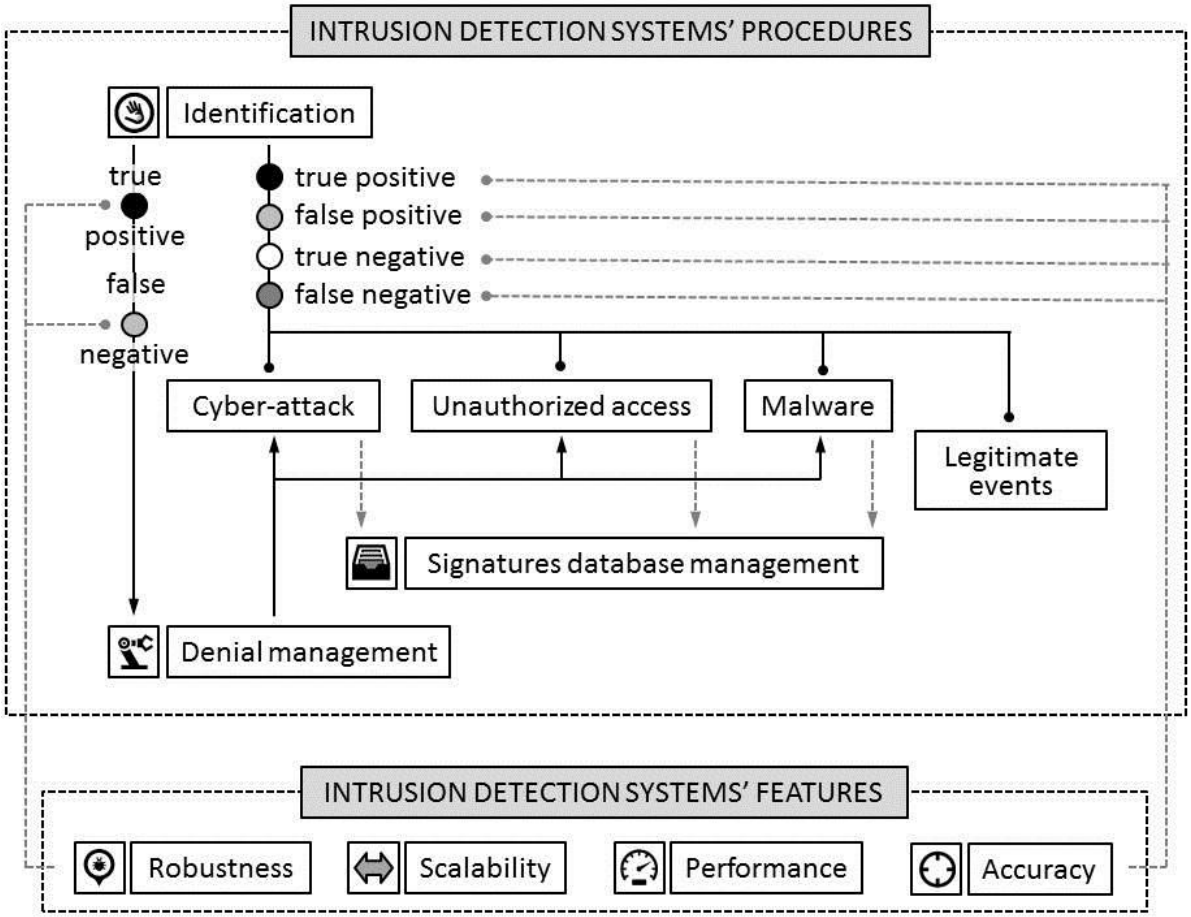

Fig. 1. Basic algorithm of intrusion prevention systems functioning and analysis

1. Classifications of intrusion detection systems

IPS development methodology includes classifications which show model of detection systems interaction with monitor environment events. There are five basic classifications [5-10] that should be discussed (Fig. 2):

- detection strategy: CA Signatures analysis, anomalies analysis, hybrid strategy;
- detection system behavior: active behavior, passive behavior;
- monitored environment: local network, global network, hybrid environment;
- detection system architecture: centralized architecture, distributed architecture, hierarchical architecture;
- detection system performance: real time analysis, offline analysis.

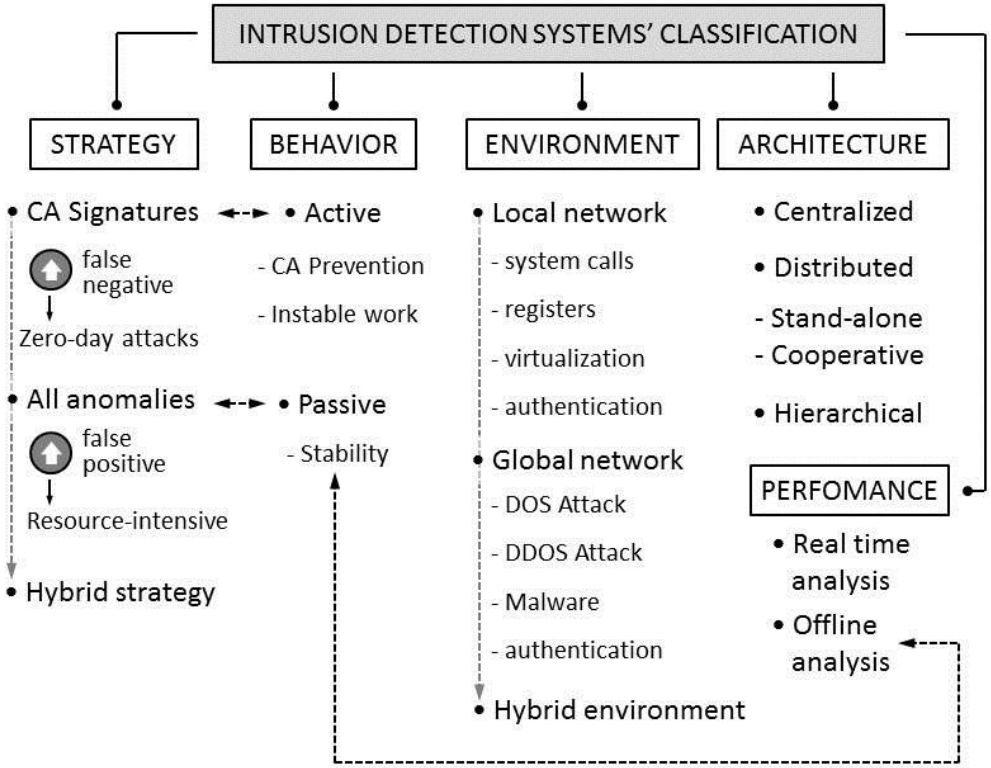

Fig. 2. Correlation between intrusion detection systems' classifications

Detection strategy depends on kind of data patterns that IPS is to identify. Originally detection systems were constructed to search for known CA signatures but nowadays due to progress of intrusion techniques it was proposed to detect and analyze all anomalies of data. Signature-based strategy is efficient against known threats but is not able to detect signatures with unknown threats, while anomaly-based strategy generate a lot of FP results, reduces IPS performance and tends to be resource-intensive. Therefore hybrid detection strategies were proposed which includes signature-based alert system block that works in real time mode and anomaly-based block for data anomalies' analysis.

The IPS behavior is determined by its allowed reaction time on the detected sample that is supposed to be cyber-threat signature. IPS which automatically provides denial management and implement countermeasures refers to active behavior system while IPS which only alerts supervisor — to passive one. Passive IPS typically has a slow react on intrusion but not so resource-intensive as active IPS and it ensure table work of Data Center infrastructure.

Monitored environment classification usually divides detection systems' models into two categories:

- network-based intrusion detection system (NIDS);
- host-based intrusion detection system (HIDS).

NIDSs are used for global network environment monitoring and HIDSs are for local environment. Moreover, there are hybrid models that combine advantages of NID and HIDS and could

be used for IPS with distributed architecture.

The IPS architecture has to be chosen up to type of monitored environment system. There are three types of architecture: centralized, distributed and hierarchical. The IPS with centralized architecture has to be composed from a single node, while the IPS with distributed architecture has to be s composed of various nodes spread at Data Center infrastructure monitored environment, so its development is more complicated because it is necessary to organize communication protocols between the different components of the detection system.

Detection system performance parameter indicates patterns analysis rate. It's obvious that data processing can be performed in real time or in offline mode. Real time detection responds to the cyber-threats before they cause major damage but to improve accuracy parameter it's necessary to combine this block with block of online analysis which will work with wider variety of threats and effective against zero-day CA.

2. Architectures of intrusion detection systems

The basic scheme of the intrusion detection framework architecture is shown at Fig. 3. It should be noticed that due to virtualization of modern Data Center platforms functional node of the scheme must not be considered as physical elements [11].

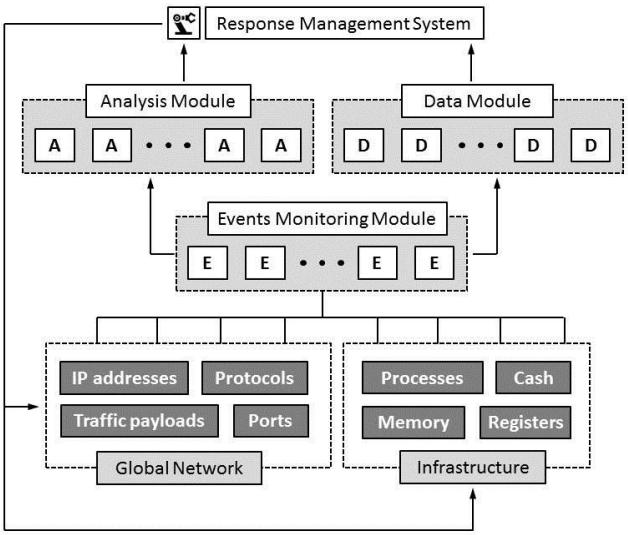

Fig. 3. Basic scheme of the intrusion detection framework architecture

The scheme includes further functional nodes (Fig. 4):

- Monitored environment: global network and Data Center infrastructure;

- Events monitoring module: E-blocks;
- Analysis module: A-blocks;
- Data module: D-blocks;
- Response management system.

As it was mentioned before, monitored environment includes global network where IP-addresses, ports, network protocols, and traffic payloads are to analyze and Data Center infrastructure local events where shared data storage, RAM, cache-memory addresses and registers are to analyze. Blocks of events monitoring module extract and collect information from the monitored environment. Analysis blocks are used for processing of collected data and detect potential cyber-threats, while data blocks assist them by storing obtained CA signatures. Response management system finally compares analysis module data with data module database and forms preventive measures interacting with monitored environment.

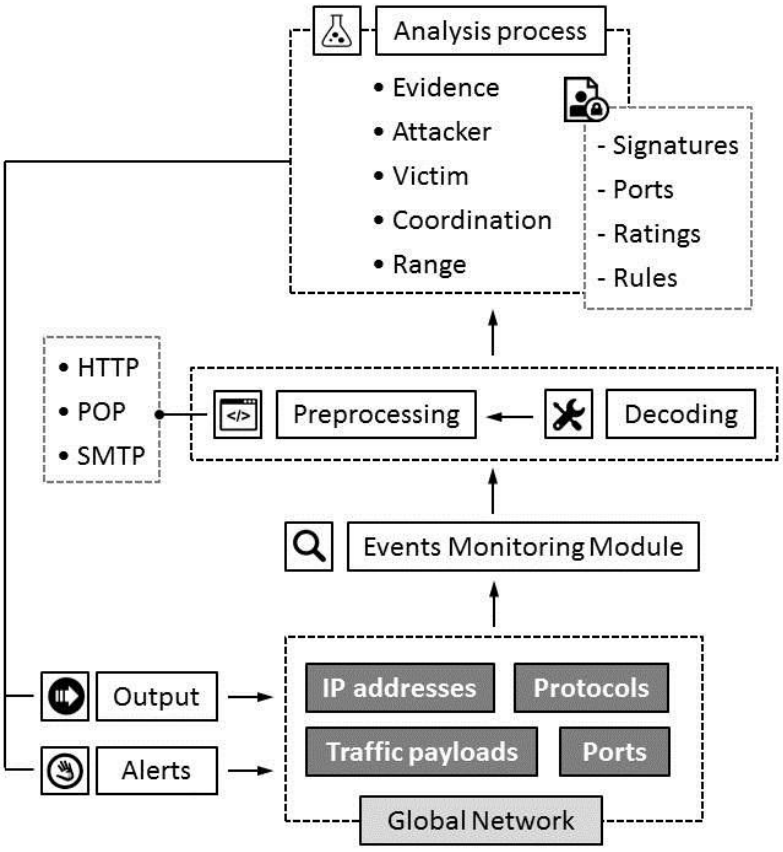

Fig. 4. Unified scheme of the intrusion detection hierarchical architecture

Modern IPS development is usually based on hierarchical architecture models. Hierarchical architecture should be considered as extension of distributed architectures which functions on cooperative mode [12] and hereby detection system is organized as set of nodes which interacts and share data. Each level of the architecture performs the pre-processing of the alerts and intensify analysis of the event. Growth of processing levels number increases accuracy and scalability of the system

but also makes it resource-intensive. Common hierarchical architecture model includes preprocessing modules which analyze the network traffic, prepare patterns of data, detect CA signatures, protect network protocols and form final alerts (Fig. 4).

3. Hybrid intrusion detection strategy

Selection of detection strategy is a key issue of IPS development methodology. It was mentioned that CA signature based detection is trivial task of hybrid detection strategy approaches, so it's more important to develop anomaly-based block which deals with the unknown cyber-threats.

Anomaly-based systems (ABS) development has to be supervised by operators and adapted to the parameters of the Data Center network while otherwise it would generate high rates of FP errors. Unlike signature based detection ABS is often considered as a black box and couldn't be classified precisely [2, 15]. Most common classification includes further groups (Fig. 5):

- statistical modeling;
- knowledge based modeling;
- modeling based on machine learning techniques.

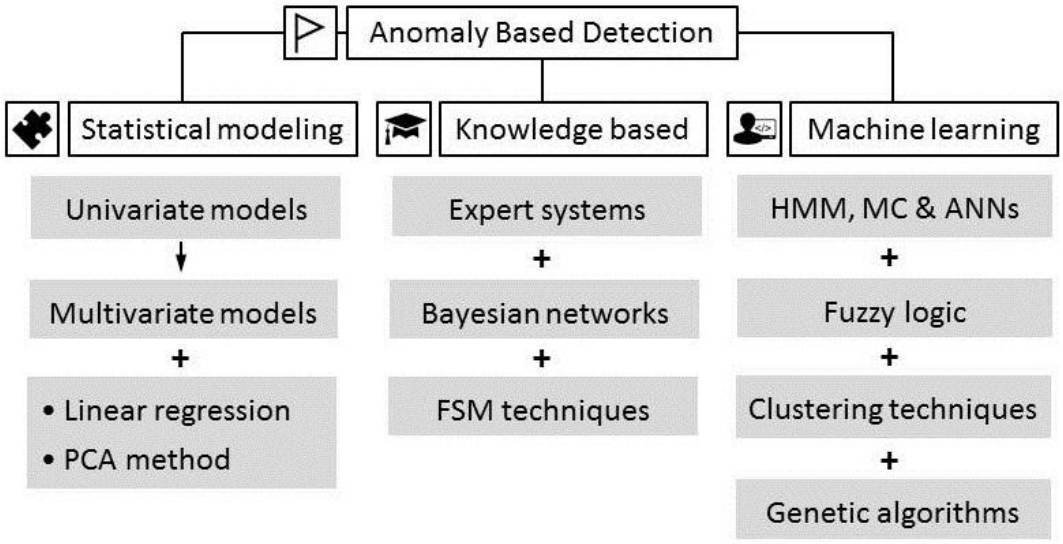

Fig. 5. Anomaly based detection systems' classification

Statistical modeling implies statistical analysis of monitored environment events and building stochastic model of legitimate processes and intrusion algorithms behavior for determination of cyber-threat probability. Basic statistical model univariate model was based on independent Gaussian random variables. Multivariate models are more preferable for IPS. They could use various metrics

and thus proved to be more scalable, adaptive and accurate.

Knowledge based modeling ABSs include expert system training stage which implies that cyber-threat detection rules should be formed directly after identification of most representative parameters of legitimate and malicious patterns database. Thereby knowledge based modeling must imply distinction between the training and modeling stages.

Most common models based on machine learning techniques are artificial neural networks (ANN) models and their predecessor, such as Hidden Markov model (HMM) or Markov chains (MC). But nowadays it was also shown that fuzzy logic application could be highly efficient and to detect the cyber threats was built methodology that interprets network traffic as fuzzy variables. Clustering techniques, in other hand, use mechanism that considers traffic samples which does not fit any of clusters as abnormal. Genetic algorithms imply support function of forming classification rules and determination model parameters.

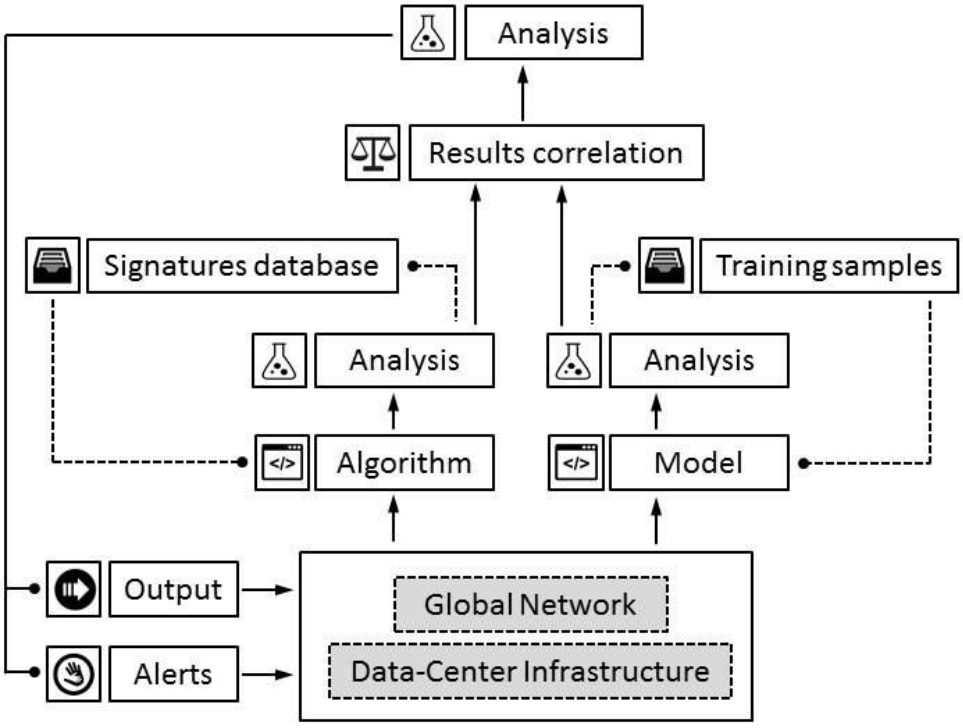

Fig. 6. Scheme of hybrid intrusion detection system

Fig. 6 shows work of combined platforms which includes ABS and signature-based detection system. It takes advantages of both system and results of detection are usually better than the results obtained by applying the strategies separately.

4. Mathematical model of intrusion detection system work

The analysis of intrusion algorithms has shown that most of

cyber-threats could be modeled as process of transmission of data in hidden channel that change state of some functional node of Data Center, such as CPU load or cache-memory registers. Thus detection system should record and analyze all changes in the system state even if the indications do not go beyond the limits defined by the Data Center security policies [1-3, 15].

Unified mathematical model of intrusion detection system work includes states of the infrastructure functional nodes, events involved in a system and transition between the states caused by those events. For binary states of one functional node $\{q_0, q_1\}$ which potentially could be used as hidden channel for illegal transmission of data we can define further set of parameters:

- $Q = \{q, q_0, q_1, q_E\}$ is a states set which includes all functional node properties
- $q \in Q$ is an initial state
- q_E is an error state (value cannot be determined);
- $\sigma = \{0,1,e\}$ is a transition functions set which simulates triggering process ($q \to q_0$, $q \to q_1$, $q \to q_E$, etc.);
- $\delta: Q \times \sigma \to Q$ is a set of transition functions.

If functional node has open access it could be triggered by signal of certain intensity or length value interval ΔI. Data is encoded in binary form so we could define ΔI_0 interval of signal as one which trigger node state to q_0 and ΔI_1 interval of signal as one which trigger state to q_1:

$$\begin{cases} I_a \le \Delta I_0 \le I_b \\ I_c \le \Delta I_1 \le I_d \end{cases}, \quad (1)$$

At the beginning of the cycle the state of node is q. If transition function σ is "0" (ΔI_0 interval of the signal) then state q will be changed to q_0, which should be expressed as $q \to q_0$. If transition function σ is "1" (ΔI_1 interval of the signal) then state q will be changed to q_1, which should be expressed as $q \to q_1$. If the interval is outside of ΔI_0 and ΔI_1 values the state q will be changed to q_E. This $q \to q_E$ transition demonstrates an error of an algorithm processing which should be determined and corrected. Error correction uses algorithm A_C which determines algorithm A_P that is based on the probability which was previously determined by HMM or MC mechanisms.

Thus, the function of switching the state of the node with open access by performing a transition function can be represented as state transition matrix (Table 1). The transition σ forms first row of the matrix, and the set $Q = \{q, q_0, q_1, q_E\}$ forms first column. Thus, the elements of the matrix are determined by the triggering of the

states which correspond to the transitions. The last line of the matrix of the transition consists algorithms A_C and A_P which deal with error state q_E.

1. State transition matrix $\delta: Q \times \sigma \to Q$.

σ	0	1	e
q	q_0	q_1	q_E
q_0	q_0	q_1	q_E
q_1	q_0	q_1	q_E
q_E	A_C, A_P	A_C, A_P	A_C, A_P

The set of finite states is $F = \{q_0, q_1\}$ which is highlighted with gray color. In the case nonbinary sequence encoding, the number of states and matrix will significantly increase, but the main principle of hidden channels detection will remain

Conclusions

Intrusion prevention systems development progresses up to modern information technologies megatrends. This evolution involves the adoption of new strategies, experience of functioning in scalable environments, organizing of hierarchical architecture and improving of detection system performance. Obtained practical results of Data Center perimeter security development demonstrate necessity of compiling intrusion prevention systems which should be based on mathematical model of detection algorithm processing.

References

1. Yeung, D.Y., Ding, Y.: Host-Based Intrusion Detection using Dynamic and Static Behavioral Models. Pattern Recognition 36/1 (2003) 229–243.

2. Undercoffer, Jeffrey L. Intrusion detection: modeling system state to detect and classify anomalous behaviors. 2004.

3. Lee, W., Miller, M., Stolfo, S.J., Fan, W.: Toward Cost-Sensitive Modeling for Intrusion Detection and Response. Journal of Computer Security10 (August 2002) 5–22.

4. Cheng, T.H., Lin, Y.D..: Evasion Techniques: Sneaking through Your Intrusion Detection/Prevention Systems. IEEE Communications Surveys Tutorials 14/4 (2012) 1011–1020.

5. Kumar, M.,: Encrypted Traffic and IPsec Challenges for Intrusion Detection System. In: Proceedings of the International Conference on Advances in Computing. (August 2012) 721–727.

6. Thonnard, O., Bilge, L., O'Gorman, G.: Industrial Espionage and Targeted Attacks: Understanding the Characteristics of an Escalating Threat. In: Proceedings of the 15th International Conference on Research in Attacks, Intrusions, and Defenses, Berlin,

Heidelberg, Springer-Verlag (2012) 64–85.

7. Wang, L., Jajodia, S., Singhal, A. K-zero Day Safety: Measuring the Security Risk of Networks Against Unknown Attacks. In: Proceedings of the 15th European Conference on Research in Computer Security, Berlin, Heidelberg, Springer-Verlag (2010) 573–587.

8. Salah, S., Maciá-Fernández, G., Díaz-Verdejo, J.E.: A Model-Based Survey of Alert Correlation Techniques. Computer Networks 57/5 (2013) 1289–1317.

9. Elshoush, H.T., Osman, I.M.: Alert Correlation in Collaborative Intelligent Intrusion Detection Systems–A Survey. Applied Soft Computing 11/7 (2011) 4349–4365.

10. Nehinbe, J.: Log Analyzer for Network Forensics and Incident Reporting. In: Proceedings of the International Conference on Intelligent Systems, Modelling and Simulation. (2010) 356–361.

11. Standard, I.: Information technology - Security Techniques - Selection, Deployment and Operations of Intrusion Detection Systems. Technical Report ISO/IEC, ISO/IEC (June 2006).

12. Gu, G., Porras, P., Yegneswaran, V., Fong, M., Lee, W.: BotHunter: Detecting Malware Infection Through IDS-driven Dialog Correlation. In: Proceedings of the 16th USENIX Security Symposium, Berkeley, CA, USA, USENIX Association (2007) 167–182.

13. Chandola, V., Banerjee, A., Kumar, V.: Anomaly Detection: A Survey. ACM Computing Surveys 41/3 (July 2009) 1–58.

14. Golovko, V., Bezobrazov, S., Kachurka, P..: Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection. Advances in Machine Learning II. Volume 263 of Studies in Computational Intelligence. Springer Berlin Heidelberg (2010) 485–513.

15. Bridges, S.M., Vaughn, R.B.: Data Mining for Intrusion Detection: From Outliers to True Intrusions. In: Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. (April 27–30 2009) 891–898.

БАЗОВІ ПІДХОДИ РОЗВИТКУ СИСТЕМ ЗАХИСТУ ЦЕНТРУ ЦЕНТРУ ОБРОБКИ ДАНИХ

А. В. Кропачьов, Д. О. Зуєв

Анотація. *Розглянуто методи кібер-захисту центрів обробки даних, що базуються на системах запобігання вторгнень на рівні внутрішньої інфраструктури і зовнішньої мережі. Проаналізовано основний алгоритм функціонування системи запобігання вторгнень і оцінки готовності, який включає в себе об'єкти аналізу, процедури і оцінку результату. Було показано, що процедури, що виконуються кібер-захистом центру обробки даних включають в себе ідентифікацію події, управління базами даних сигнатур кібератак і систему контролю. Було продемонстровано, що ефективність системи запобігання вторгнень ґрунтується на числі помилок і масштабованості системи. Таким чином, критерій ефективності повинен включати в себе точність, надійність, продуктивність і параметри*

масштабованості. Обговорювалися основні системи класифікації, які грунтуються на моделі взаємодії систем виявлення з потенційно небезпечними подіями. Зокрема, розглянута класифікація на основі стратегії виявлення, включає аналіз сигнатур кібер-атак, аналіз аномалій, гібридну стратегію, класифікація заснована на поведінці системи виявлення, яка включає в себе моделі активного та пасивного поведінки системи, класифікація на основі середовища моніторингу, яка включає в себе роботу в локальній мережі, глобальної мережі і гібридну модель, класифікація по архітектурі системи виявлення, яка включає в себе централізовану архітектуру, розподілену архітектуру та ієрархічну архітектуру, а також класифікація швидкості відгуку системи виявлення, яка включає аналіз в реальному часі і оффлайн-аналіз. Було згадано, що розробка систем на основі аналізу аномалій повинна контролюватися операторами і адаптуватися до параметрів мережі центру обробки даних. Дані системи були поділені на три групи: статистичне моделювання, моделювання, засноване на управлінні знаннями, і моделювання на основі методів машинного навчання. Було згадано, що кібер-загрози можуть бути змодельовані як процес передачі даних з прихованого каналу, які змінюють стан функціонального вузла центру обробки даних. Запропонована уніфікована математична модель роботи системи виявлення вторгнень, яка включає аналіз станів функціональних вузлів інфраструктури, подій і фактів переходу між станами.

***Ключові слова**: центр обробки даних, система запобігання вторгнень, надійність, гібридне середовище, система аналізу аномалій, машинне навчання*

БАЗОВЫЕ ПОДХОДЫ РАЗВИТИЯ СИСТЕМ ЗАЩИТЫ ЦЕНТРА ЦЕНТРА ОБРАБОТКИ ДАННЫХ

А. В. Кропачев, Д. О. Зуев

***Анотация.** Рассмотрены методы кибер-защиты центров обработки данных, которые базируются на системах предотвращения вторжений на уровне внутренней инфраструктуры и внешней сети. Проанализирован основной алгоритм функционирования системы предотвращения вторжений и оценки готовности, который включает в себя объекты анализа, процедуры и оценку результата. Было показано, что процедуры, выполняемые кибер-защитой центра обработки данных включают в себя идентификацию события, управление базами данных сигнатур кибератак и систему контроля. Было продемонстрировано, что эффективность системы предотвращения вторжений основывается на числе ошибок и масштабируемости системы. Таким образом, критерий эффективности должен включать в себя точность, надежность, производительность и параметры масштабируемости. Обсуждались основные системы классификации, которые основываются на модели взаимодействия систем обнаружения с потенциально опасными событиями.*

В частности, рассмотрена классификация на основе стратегии обнаружения, включающая анализ сигнатур кибер-атак, анализ аномалий, гибридную стратегию, классификация основанная на поведении системы обнаружения, которая включает в себя модели активного и пассивного поведения системы, классификация на основе среды мониторинга, которая включает в себя работу в локальной сеть, глобальной сети и гибридную модель, классификация по архитектуре системы обнаружения, которая включает в себя централизованную архитектуру, распределенную архитектуру и иерархическую архитектуру, а также классификация скорости отклика системы обнаружения, которая включает анализ в реальном времени и оффлайн-анализ. Было упомянуто, что разработка систем на основе анализа аномалий должна контролироваться операторами и адаптироваться к параметрам сети центра обработки данных. Данные системы были разделены на три группы: статистическое моделирование, моделирование, основанное на управлении знаниями, и моделирование на основе методов машинного обучения. Было упомянуто, что кибер-угрозы могут быть смоделированы как процесс передачи данных по скрытому каналу, которые изменяют состояние функционального узла центра обработки данных. Предложена унифицированная математическая модель работы системы обнаружения вторжений, которая включает анализ состояний функциональных узлов инфраструктуры, событий, и фактов перехода между состояниями.

Ключевые слова: *центр обработки данных, система предотвращения вторжений, надежность, гибридная среда, система анализа аномалий, машинное обучение.*