I. J. Computer Network and Information Security, 2017, 4, 10-21
Published Online April 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2017.04.02

Anomaly Detection System in Secure Cloud Computing Environment

Zhengbing Hu
School of Educational Information Technology, Central China Normal University, Wuhan, China
E-mail: hzb@mail.ccnu.edu.cn

Sergiy Gnatyuk, Oksana Koval, Viktor Gnatyuk and Serhii Bondarovets
National Aviation University, IT-Security Academic Dept, Kyiv, Ukraine
E-mail: s.gnatyuk@nau.edu.ua, oksanakoval@mail.ua, viktorgnatyuk@ukr.net, bondss29@gmail.com

Abstract—Continuous growth of using the information technologies in the modern world causes gradual accretion amounts of data that are circulating in information and telecommunication system. That creates an urgent need for the establishment of large-scale data storage and accumulation areas and generates many new threats that are not easy to detect. Task of accumulation and storing is solved by datacenters – tools, which are able to provide and automate any business process. For now, almost all service providers use quite promising technology of building datacenters – Cloud Computing, which has some advantages over its traditional opponents. Nevertheless, problem of the provider's data protection is so huge that risk to lose all your data in the "cloud" is almost constant. It causes the necessity of processing great amounts of data in real-time and quick notification of possible threats. Therefore, it is reasonable to implement in data centers' network an intellectual system, which will be able to process large datasets and detect possible breaches. Usual threat detection methods are based on signature methods, the main idea of which is comparing the incoming traffic with databases of known threats. However, such methods are becoming ineffective, when the threat is new and it has not been added to database yet. In that case, it is more preferable to use intellectual methods that are capable of tracking any unusual activity in specific system – anomaly detection methods. However, signature module will detect known threats faster, so it is logical to include it in the system too. Big Data methods and tools (e.g. distributed file system, parallel computing on many servers) will provide the speed of such system and allow to process data dynamically. This paper is aimed to demonstrate developed anomaly detection system in secure cloud computing environment, show its theoretical description and conduct appropriate simulation. The result demonstrate that the developed system provides the high percentage (>90%) of anomaly detection in secure cloud computing environment.

Index Terms—Anomaly Detection, Big Data, Information Security, Data Analysis, Machine Learning, Signature Detection, Data Center, Cloud Computing, Vulnerability, Security, Technology Architecture, Threat Model.

I. Introduction

Anomaly detection is one of the most important concepts of data analysis. Information object is considered as an anomaly if it is significantly differs from normal data behavior in some sphere. In general, it means that object is not like the others in a particular data array [1]. It is important to detect these objects in order to consider them from a different angle and use other detection methods. During the anomaly detection process researchers deal with such problems: as determing of normal area that might be presented in adequate form is often a difficult task; boundary between normal and anomaly behavior is not always clear; exact anomaly detection is different depending on field of application; availability of relevant data for training or checks; data can contain noise; normal behavior is dynamic and constantly evolving.

Anomaly detection methods are widely used in the following areas: cloud-computing environment, fraud detection in banking and mobile areas, monitoring of information systems hardware, network's intrusions detection system, processing CCTV images, detection of suspicious web-site etc.

From this point of view the aim of this paper is to develop an anomaly detection system in secure cloud computing environment. To achieve this aim should be solved such tasks:

- Developing of secure cloud data center model;
- Developing of anomaly detection system for Cloud Computing protected environment;
- Big Data concept analysis;
- Experimental research of anomaly detection module in developed system for Cloud Computing secure environment.

II. The Analysis of Existing Research and Problem Definition

In order to solve problems described in Section I it is needed to analyze next issues: modern type of data centers; Cloud Computing technology; modern data centers models; Big Data conception; anomaly detection methods.

A. Modern Type of Data Centers

Today data centers provide large number of services, which specific depends on the data center type. There are several types of data centers [25, 27, 29]: private cloud providers; scientific computing centers; co-location data centers; in-house data centers; wholesale data centers; dedicated hosting; shared hosting; managed hosting.

Important data center characteristic is a set of components, namely Tier, which is an attribute of what it can offer to customer, for example, physical infrastructure, cooling system, supply system and expected uptime level. All these characteristics define the redundancy of all infrastructures. There are four data center tiers, each of which includes previous and has bigger uptime level [10, 12].

Co-operation between data center components depends on its architecture. Typical architecture includes utility system, security system, IT-infrastructure and monitoring system, that controls other system [24].

General specification of data center tiers is shown in Table 1.

Table 1. Data Center Tiers

Tier 1	Non-redundant capacity components (single uplink and servers)
Tier 2	Tier 1 + Redundant capacity components
Tier 3	Tier 1 + Tier 2 + Dual-powered equipments and multiple uplinks
Tier 4	Tier 1 + Tier 2 + Tier 3 + all components are fully fault-tolerant including uplinks, storage, chillers, HVAC systems, servers etc. Everything is dual-powered

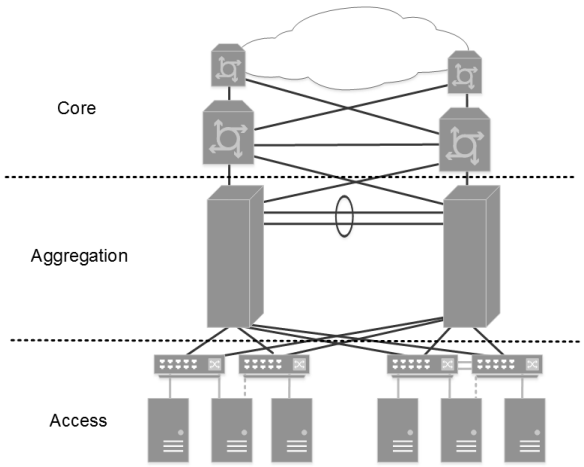


Fig.1. Basic Layered Design

Necessary aspect of data center design is using a multilevel approach is a basic aspect of data center design, because it improves scalability, performance, flexibility, resiliency, and maintenance [8]. Designing a flexible architecture that has the ability to support new applications in a short time can lead to a significant competitive advantage. Such design requires solid initial planning and thoughtful consideration in the areas of port density, access layer uplink bandwidth, true server capacity, and oversubscription [9]. Fig. 1 shows the basic layered design.

B. Cloud Computing Technology

In fact, Cloud Computing is a providing on-demand computing resources (all from applications to data centers) to customers over Internet-based payment [5]. Cloud Computing exhibits the following key characteristics [23]: self-service on-demand; wide network access; resource pooling; rapid elasticity; measured service.

The NIST's definition of cloud computing defines the service models as follows [26].

Software as a Service (SaaS). The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure [18].

Platform as a Service (PaaS). The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider [16].

Infrastructure as a Service (IaaS). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

Moreover, there are various cloud computing deployment models:

1) Private Cloud is infrastructure operated solely for a single organization, whether managed internally or by a third-party, and hosted either internally or externally [20];
2) Public Cloud when the services are rendered over a network that is open for public use[6];
3) Hybrid Cloud is a composition of two or more clouds (private, community or public) that remain distinct entities but are bound together, offering the benefits of multiple deployment models [13].

Requirements for information security in "cloud" data centers based on the reference model architecture that is described in "Security Recommendations for Cloud Computing Providers (CPS)" by Federal Agency for German Information Security (BSI). This reference architecture (Fig. 2) approximately indicates components common to many cloud computing platforms [11, 17, 19, 28].

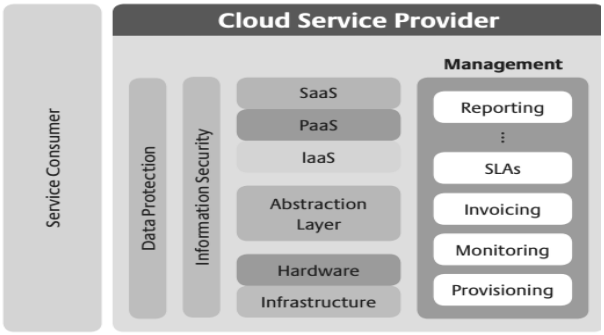


Fig.2. Reference Architecture for Cloud Computing Platform

C. Modern Data Center Models

Table 2 shows comparative analysis of data centers models in terms of architecture and information security. Symbolic notation: "+" – match the criteria; "–" no compliance with the criteria; "?" – lack of information in open sources; *C* – confidentiality; *I* – integrity; *A* – availability.

All studied data centers in table 2 are built on one or combine several technologies (services of models). Set of components corresponds to highest Tier levels, what means that infrastructures are almost or completely failsafe, support systems have backup components, and the expected uptime more than 99%.

There were analyzed monitoring and ensuring data protection system for each data center and how it is shown in Table 2 only three data centers have reliable and official data related to information security (BEMOBILE, Tulip Data Center and Switch Super NAP).

Table 2. Analysis of the Known Implemented Data Centers Models

Data center model	Technology	Set of components	C	I	A	Lack of vulnerabilities
Volia Data Center	SaaS	Tier 2/Tier 3	+	?	+	–
Data Center DataGroup	SaaS	Tier 3	?	+	+	–
BEMOBILE	SaaS	Tier 3 (communication Tier 4)	+	+	+	–
Amazon Data Center	SaaS, PaaS, IaaS	Tier 4	+	?	+	–
Google Data Center	SaaS, PaaS, IaaS	Tier 3+	?	?	?	–
Yandex Data Center	SaaS	Tier 3	+	?	+	?
Tulip Data Center	IaaS	Tier 3+	+	+	+	+
Lakeside Tech. Center	IaaS	Tier 4	?	+	+	?
Microsoft Data Center	SaaS, PaaS	Tier 4	+	?	+	–
Range International IG	PaaS, IaaS	Tier 4	?	?	+	?
Switch Super NAP	SaaS, IaaS	Tier 3/ Tier 4	+	+	+	?
DuPont Fabros Tech.	SaaS, PaaS, IaaS	Tier 4	+	?	+	+
Utah Data Center	SaaS	Tier 3/ Tier 4	?	?	?	–

The absence of known vulnerabilities is also an important criterion [7, 24]. According to Table 2 almost for all data centers there were recorded different, from powerful lightning strike to the building of the data center or multiple network attacks. The only data centers for which vulnerabilities were not detected (or information about them is hidden) – are Tulip Data Center and DuPont Fabros Technology.

D. Anomaly Detection Methods

The analysis of modern anomaly detection methods allowed to make their comparison (table 3) by following criteria [3]:

- Low demand on computing resources (LDCR);
- Lack of need in particular data distribution (LNDD);
- Simplicity of implementation (SI);
- Little amount of false-positive rate (LAFPR);
- Unsupervised learning (UL).

According to the analysis, Decision Tree method is one of the best bases for developing anomaly detection system.

However, drawbacks uniting all of the above methods are following:

- Unprotected state of the information system, while anomaly detection system is learning and building normal profile;
- If malicious activity corresponds to normal profile, there will be no alert about anomaly;
- High false-positive rate;
- Notifications and warnings about anomalies can contain not enough information for the further analysis because of aggregation of big amount of data and abstraction from particular information for moving to mathematical modeling [2].

Signature databases did not manage to update intime, that's why we propose to use a system, which combines detection of new anomalies and tracking existing, using signature methods and available databases. To increase the speed of such system, it is recommended to use Big Data methods and instruments.

Table 3. Multicriterial Analysis of Modern Anomaly Detection Methods

Method	Criteria				
	LDCR	LNDD	SI	LAFPR	UL
Neural networks	+	+	+/–	+	–
Bayesian networks	+	+	+/–	+	–
Support Vector Machine	+	+	+/–	+	–
Decision Tree	+	+	+	+	–
K-nearest neighbor	–	+/–	+	+/–	+
Relative density	–	+/–	+	+/–	+
Clusterization	+/–	+	+	–	+
Parametric methods	+	–	–	+/–	+
Non-parametric methods	+	–	–	+/–	+
Kolmogorov complexity	+/–	+	+/–	+/–	+
Entropy	+/–	+	+/–	+/–	+
PCA	–	+/–	+/–	+/–	+

III. The Proposed Solution

The research proposes an anomaly detection system in secure cloud computing environment. Designed system contain of secure cloud data center model in which there was implement anomaly detection system based on Big Data concept.

A. Development of the secure data center model based on Cloud Computing

The first stage of model development is using technological architecture that includes three main "building" blocks.

1) 10 Gigabit Ethernet

A cloud data center is designed with the high density of virtual machines coupled with a high processor core count. From a networking perspective, the increase in virtual machine and processor core density promotes a transition to 10 Gigabit Ethernet as the required mechanism for attaching servers. Specific benefits include: real-time policy-based configuration; mobile security and network policy; nondisruptive management model, aligning management and operations environments for virtual machines and physical server connectivity in the data center.

2) Unified Fabric

This block gives all servers (physical and virtual) access to the LAN, SAN, and IPC networks, allowing more to be consolidated in the customer's network for greater efficiency and costs savings.

3) Unified Computing

It enables a fully virtualized cloud data center with pools of computing, network, and storage resources. The Unified Computing bridges the silos in the classic data center, enabling better utilization of infrastructure in a fully virtualized environment, and creates a unified architecture using industry-standard technologies that provide interoperability and investment protection.

Fig. 3 shows technological architecture, which represent next generation data center based on cloud computing. The diagram shows only examples of blocks for the data center. In total over architecture includes not only structure components but also it is governed by different types of service and regulatory requirements.

There are 9 network layers in architecture (Fig. 3): application software; virtual machine, VSwitch; storage, SAN; compute; access; aggregation; core; peering; IP-NGN backbone.

Each layer is connected to the previous with a specific connection type. From application software layer to Virtual machine & VSwitch layer there is App to HW/VM connection type. Than application data come to distributed virtual switches VSwitch.

After that data from SAN and application data from VSwitch transfer to computing layer via 4G FC (fibrechanel) and VSwitch to HW. Computing result transfers to access layer via 4G FC, 10G FCoE (FibreChanneloverEthernet) and 1G Ethernet, and after that data is transmitted to aggregation layer via 10G Ethernet. On this layer it is possible to control app, services and establish firewall services (IDS, SSL, anti-DDoS).

The next layer is core, where procedures of global positioning and intrusion detection are also applying. Peering layer is responsible for secure domain routing. The last layer is Internet, where 10G Ethernet connection is used.

Along with the technological architecture of data centers an important place also occupies question of confidence in cloud computing infrastructure model. Fig. 4 shows structure of secure cloud data center from the perspective of security, for example threat model and measures to be taken to minimize risks. Structure also represents full control, compliance and SLA.

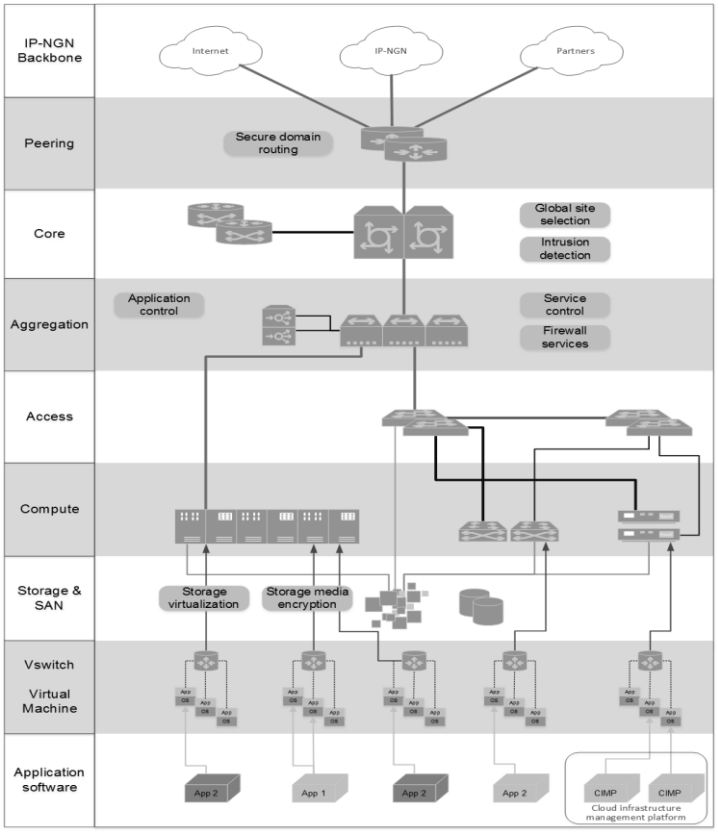


Fig.3. Technology Architecture of Data Center Based on Cloud Computing

The main idea of this model is that information security should not be secondary part of overall security. It must be applied and implemented at all levels of architecture.

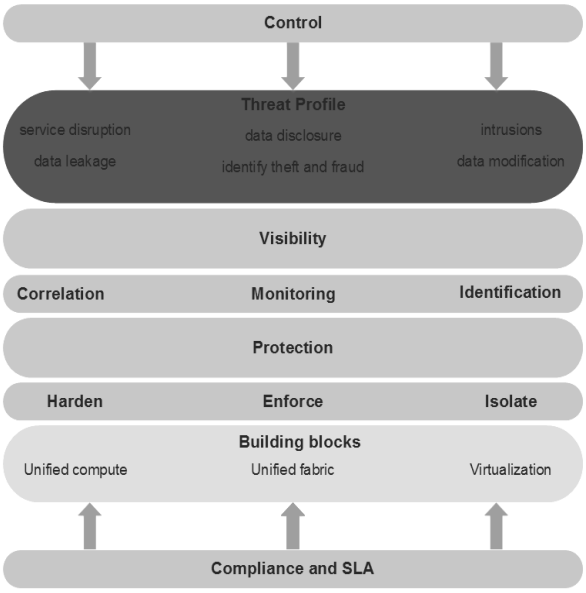


Fig.4. Structure of Secure Data Center Based on Cloud Computing

Construction of secure cloud data center architecture includes the implementation of six levels security: 1) Physical Protection; 2) Server Protection; 3) Data Protection; 4) Protection of Application and Platforms; 5) Network Security; 6) Secure Encryption system and Key Management System.

Fig. 5 shows relation between levels of cloud data center protection and their interaction.

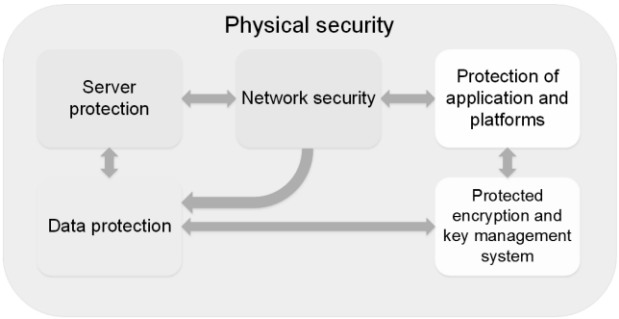


Fig.5. Levels of Secure Cloud Data Center Architecture from the Perspective of Information Security

B. Anomaly detection system based on Big Data

It is proposed to implement anomaly detection system based on Big Data concept using data center resources. Common structure of such system is shown on fig.6.

Input data arrives at two modules in parallel, then Master Node start working in each of them, distributing load between Slaves, where two-step MapReduce method is implemented. The output is useful data, which is checking by conditions and resulting in either normal data, or classified threat, or unknown activity.

Hybrid system logic is presented in Table 4.

Table 4. Common Hybrid System Logic

	Anomaly Detection	**Misuse Detection**	Explanation
Check	0	0	Normal data
	1	0	Threat detected
	0	1	Threat detected and classified
	1	1	Threat detected and classified

As Misuse Detection module it is recommended to use open-source application Snort, which works on both Windows and Linux operating system.

Snort – is intrusion detection system (IDS), which is an extremely powerful tool, even compared with commercial IDS. Many users share their security rules in Snort community, what is useful when it is necessary to have the most recent rules.

Snort can be used in 4 modes:

- Sniffer mode – reading of network's traffic and displaying it on the screen;
- Packet logger mode – writing network's traffic in file;
- IDS mode – network's traffic which is corresponding to rule is written;
- IPS mode – modified version of previous mode. It accepts packets from firewall, compares them with signature rule and marks them as "Discard" if they respond the rule [22].

We use Decision Tree Method as Anomaly Detection module, which is based on building "decision tree". That tree contains nodes (internal and terminal) and branches.

Internal nodes are the ones that split into two children. Each internal node corresponds to one of the input features; there are edges to children for each of the possible values of that input feature.

A terminal node has a class label associated with it, such as observations that fall into the particular terminal node are assigned to that class. To use a decision tree, a feature vector is presented to the tree. If the value for a feature is less than a defined number, then the decision is moves to the left child. Otherwise – moves to the right one.

Process continues, until it reaches one of the terminal nodes and the class label that corresponds to the terminal node is the one that is assigned to the pattern.

Decision tree induction algorithms are functioning recursively:

- First, a feature must be selected as a root node;
- In order to create the most efficient (the smallest) tree, the root node must effectively split the data. Each split attempts to pare down a set of instances (the actual data) until they all have the same classification. The best split is the one that provides what is termed the most information gain.

The tree grows by recursively splitting each node using the feature which gives the best information gain until the leaf is consistent.

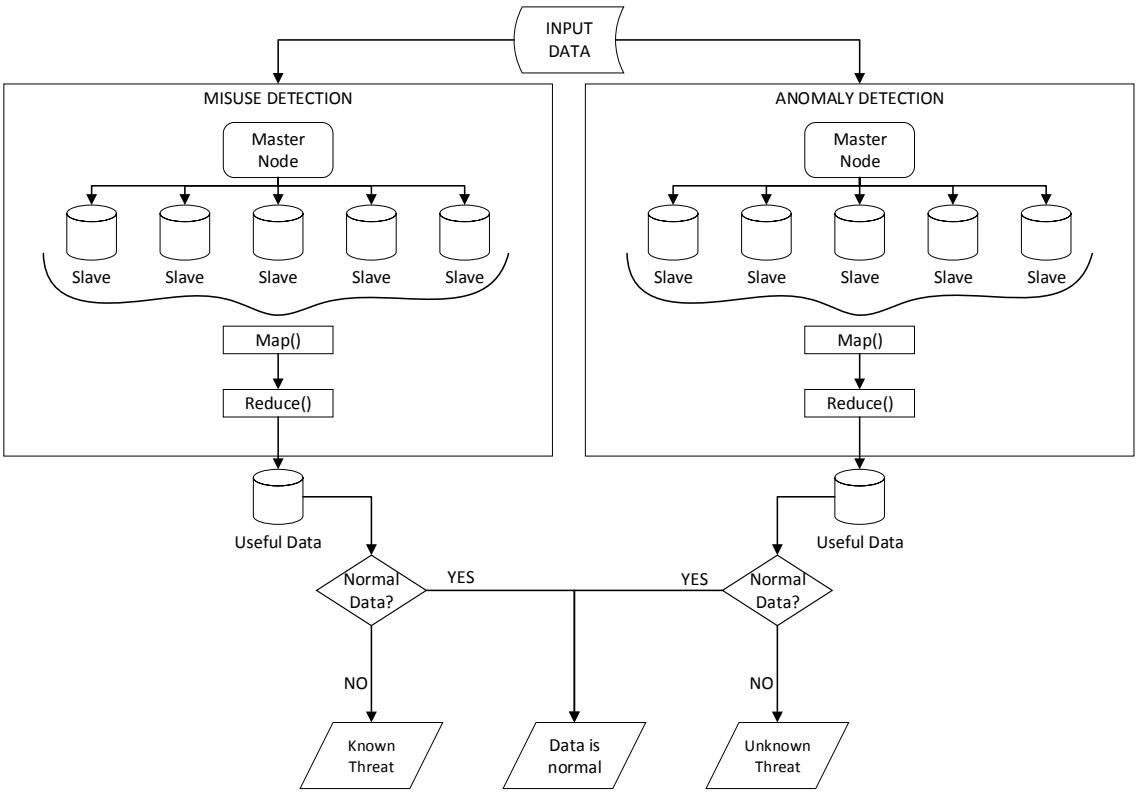


Fig.6. Hybrid Anomaly Detection System Structure

Four next steps are used to calculate information gain:

1. Calculate entropy at node A (Fig.7):

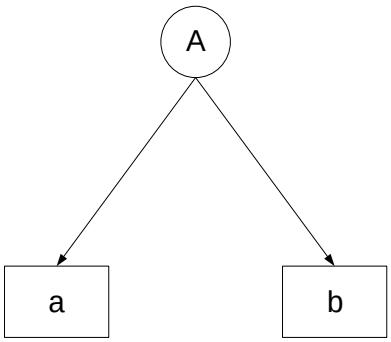


Fig.7. Model Example

$$H(S) = -\left(\frac{M}{M+N}\right)\times\log_2\left(\frac{M}{M+N}\right) - \left(\frac{N}{N+M}\right)\times\log_2\left(\frac{N}{N+M}\right),$$

where M – quantity of anomaly data in the node A, N – quantity of normal data in the node A, $H(S)$ - value of entropy before the split.

2. The data set is split into two branches by different feature; the entropy for each branch is calculated:

$$H_a = H(m,n);$$

$$H_a = -\left(\frac{m}{m+n}\right)\times\log_2\left(\frac{m}{m+n}\right) - \left(\frac{n}{n+m}\right)\times\log_2\left(\frac{n}{n+m}\right),$$

$$H_b = H(M-m, N-n),$$

$$H_b = -\left(\frac{M-m}{(M-m)+(N-n)}\right)\times\log_2\left(\frac{M-m}{(M-m)+(N-n)}\right) - \left(\frac{N-n}{(N-n)+(M-m)}\right)\times\log_2\left(\frac{N-n}{(N-n)+(M-m)}\right),$$

where m – quantity of anomaly data in the node a, n – quantity of normal data in the node a.

3. The entropy for each branch is added proportionally to get total entropy for the split:

$$H(S \mid A) = P_a \times H_a + P_b \times H_b,$$

$$H(S \mid A) = \left(\frac{m+n}{M+N}\right)\times H_a + \left(\frac{(M-m)+(N-n)}{(M+N)}\right)\times H_b$$

where P_a – ratio between the quantity of node's a elements and the quantity of node's A elements, P_b – ratio between the quantity of node's b elements and the quantity of node's A elements.

4. The resulting entropy is subtracted from the entropy before the split and the result is the information gain or decrease in entropy:

$$I.G.(S, A) = H(S) - H(S \mid A).$$

Decision tree is a greedy algorithm that grows the tree top-down. At each node it selects the features that best classifies the local training samples. This process continues until the tree perfectly classifies the training samples, or all features have been used [2].

C. Big Data concept

To process big amounts of data, a set of special methods is used. One of the examples is MapReduce [1].

MapReduce – is software framework for distributed computing, which uses "divide and conquer" method for splitting big data's difficult problems into the small blocks of work and processing them in parallel mode.

MapReduce contains two steps: step "Map" – data from the master node splitting into great amount smaller subproblems. Worker nodes process some subsets under the JobTracker's control and save the result in the local file system. Step "Reduce" - analyses and perform operation of merge the input data from the previous step. A large number of Reduce-step is possible in order to execute processes of merge in parallel mode, so these tasks are also performed on worker nodes under the JobTracker's control.

Another method is Hadoop. Hadoop contains distributed file system; platforms for data analysis and storage; parallel computing management level; configurations administrations.

One more utility is Apache Spark. Spark – is cluster-computing engine, which provides extremely fast data processing and reliability. It has software interface, which are based on different programming languages: Java, Python, and Scala.

It supports in-memory computing, which allows access to data and process requests much faster, compared to disk-based system (i.e. Hadoop).

In general, Spark is progressive and very useful update for Hadoop, aimed at improvement of real-time analysis.

The main advantages of Apache Spark:

- The fastest engine for processing big arrays of data;
- Worker processes are identified using MapReduce-style, which simplifies its implementation along with Hadoop;
- Simple installation;
- Spark is written in Scala, modern object-oriented programming language, which has many resources and active community;
- Many platforms is supporting Spark and its technology stack (MapR, Cloudera, Databricks);
- Spark's reliability can be proved by Intel recommendation to use it in healthcare solutions;
- One of the most used Spark features – capability to consolidate data sets from a few incompatible sources [8].

IV. Simulation and Result

In this section, several experiments are carried out on the developed system to check the security level and check the classification effectiveness of chosen anomaly detection method.

A. Experiment 1

The aim of the experiment: check the security level of cloud data center.

CloudSim simulation system

CloudSim platform is a generic and scalable simulation tool that allows a complete modeling and simulation of cloud computing systems and infrastructures, including the construction of cloud data centers. It is an extension of the basic functionality GridSim platform, enabling simulation data stores, web services, the distribution of resources among virtual machines.

The model of secure data center is implemented on CloudSim platforms as follows:

1) Set up of service and Internet provider for data storage and using of cloud services in cloud computing environment;
2) Launch of time analysis and resource usage module;
3) Use of heuristic algorithm for task scheduling and real-time modeling;
4) Effective provision of resources and measure productivity on the basis of the algorithm;
5) Cloudsim use Green Computing, which allows to achieve energy efficiency and power utilization;
6) Important place is taken by hypervisor security in the cloud;
7) Including of cloud computing security modules by using simulation tools for distributed denial attack infrastructure and impact analysis tool for DDoS attacks;
8) Use of proposed in 2.1 hierarchical data center model by phased connection of different layers of architecture;
9) Application of security policies from protected virtual machines policy to location monitoring system.

Table 5 shows characteristics of designed data center model based on cloud computing from the perspective of architecture.

Table 5. Characteristics of Designed "DCM" Data Center Model Based on Cloud Computing

Data-center	Technology	Set of components	Cloud model
DCM	IaaS, SaaS	Tier 3	Гібридна

Fig. 8 shows a diagram of data center simulation on CloudSim platform, which is realize in real-time mode.

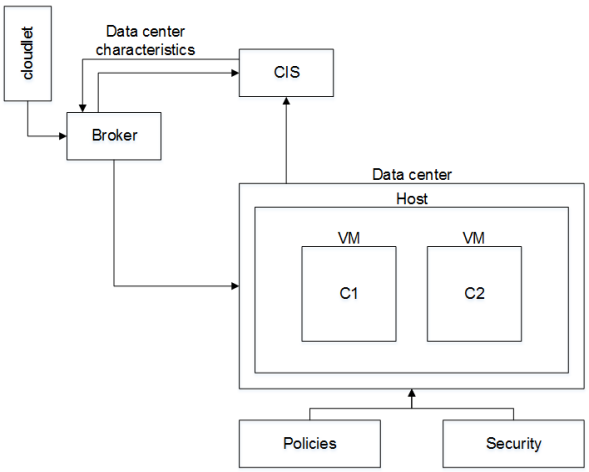


Fig.8. Scheme of Modeling Process

A complete model of secured data center based on Cloud Computing is constructed by connection the base data center model with different levels of protection, for example: protection against malicious software; configured firewall; remote management using SSH, TLS / SSL, IPSec; multifactor authentication; use of regular backup copying; use of virtual machines policies and more.

There were conducted three experiments to study the developed system (Fig. 9). Comparison results of simulations on a platform CloudSim is given in Table 6:

Table 6. Comparing the results of the simulations

№	Security levels	Efficiency	Detected and neutralized attacks
1	100%	99,89%	99,65%
2	<50%	67,23%	45,78%
3	0%	0,0%	0,0%

The experimental results indicate that while connected all levels of protection efficiency (which refers to the "performance + data security") and the level of detected and neutralized attacks is almost 100%; during the second experiment with at least 50% of connected layers of security shows that efficiency and detected and neutralized attacks dropped almost in half; and during the third experiment, when the level of protection does not connected, the efficiency and the level of detected and neutralized attack is zero.

OPNET IT simulation system

OPNET IT (Riverbed Modeler) platform is tool for creating and modeling of infrastructures scripts using cloud computing. The designed model was implemented using this simulation system from the perspective of technological architecture, with connecting some built-in protection components: the "cloud" servers, client.

Simplified technological model carried over to the platform OPNET IT is shown in Fig. 10.

Using the built-in properties of the components of OPNET IT lets connect security levels. Unfortunately, not all security levels see 2.1 can be included, given the specific platform OPNET IT.


```
========== OUTPUT ==========
Cloudlet ID    STATUS    Data center ID    VM ID    Time    Start
    1        SUCCESS         2              1        80       0,1
    0        SUCCESS         2              0        160       0,1

Architecture:

Simulation of secure cloud data center

Using service model: SaaS, IaaS
Ecnryption and key managemetn - ON
Security measures against malware - ON
Configured firewall - ON
Remote administration - possible. Using SSH, TLS/SSL, IPSec
Multi-factor authentification - ON
Regular data backups - ON
Rights Management - Least Privilege Model
Logging and monitoring of data center - ON

Data exchange started.............
Perfoming task started.............
.....................................
Data exchange completed.............
Perfoming task completed.............
Perfoming task completed.............
Data status - secure

Efficiency -  99.89%
Detected and neutralized threats and incidetns -  99.65%
```

a

```
========== OUTPUT ==========
Cloudlet ID    STATUS    Data center ID    VM ID    Time    Start Time
    1        SUCCESS         2              1        80       0,1
    0        SUCCESS         2              0        160       0,1

Architecture:

Simulation of secure cloud data center

Using service model: SaaS, IaaS
Ecnryption and key managemetn - ON
Security measures against malware - OFF
Configured firewall - ON
Remote administration - possible. Using SSH, TLS/SSL, IPSec
Multi-factor authentification - ON
Regular data backups - OFF
Rights Management - Least Privilege Model
Logging and monitoring of data center - OFF

Data exchange started.............
Perfoming task started.............
.....................................
Data exchange completed.............
Perfoming task completed.............
Perfoming task completed.............
Data status - security problem

Efficiency -  67.23%
Detected and neutralized threats and incidetns -  45.78%
```

b

```
========== OUTPUT ==========
Cloudlet ID    STATUS    Data center ID    VM ID    Time    Start Time
    1            0
Architecture:

Simulation of secure cloud data center

Using service model: SaaS, IaaS
Ecnryption and key managemetn - OFF
Security measures against malware - OFF
Configured firewall - ON
Remote administration - none
Multi-factor authentification - ON
Regular data backups - OFF
Logging and monitoring of data center - OFF

Data exchange started.............
Perfoming task started.............
.....................................
Data exchange ... error occured.............
Data exchange completed.............
Perfoming task completed.............

Efficiency -  0.0%
Detected and neutralized threats and incidetns -  0.0%
```

c

Fig.9. Connection to the Base Model, All Security Levels (a) Random Layers of Security (b) Without Using Any Security Levels (c)

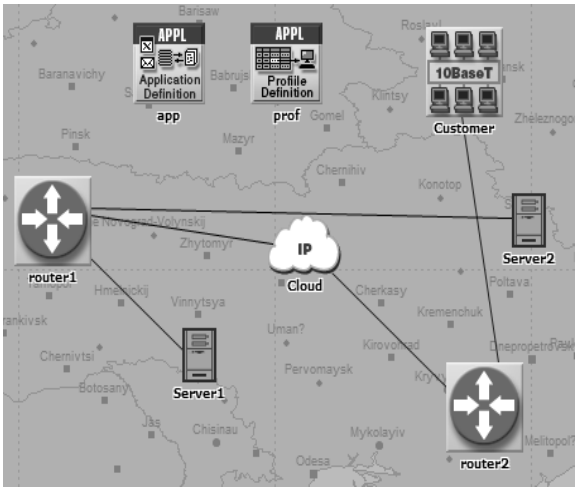


Fig.10. Adapted to the Simulation Environment Model of a Secure Data Center

The next step is configuration of simulation process; determine the time during which it will occur, and actual launch. When you run a simulation it adapts to real time (e.g. 1:00 the simulation platform OPNET IT really will be 1 min.). Log Files with simulation results reflect only general information, progress, speed, time (Fig. 11).

```
156  | Speed: Average (827,711 events/sec.); Current (712,253 events/sec.)
157  | Time : Elapsed (6.6 sec.); Remaining (38 sec.)
158  | DES Log: 8 entries
159  |-------------------------------------------------------------------
160  |-------------------------------------------------------------------
161  | Progress: Time (11 min. 11 sec.); Events (6,000,143)
162  | Speed: Average (828,176 events/sec.); Current (833,334 events/sec.)
163  | Time : Elapsed (7.2 sec.); Remaining (35 sec.)
164  | DES Log: 8 entries
165  |-------------------------------------------------------------------
166  |-------------------------------------------------------------------
167  | Progress: Time (11 min. 59 sec.); Events (6,500,148)
168  | Speed: Average (833,458 events/sec.); Current (902,534 events/sec.)
169  | Time : Elapsed (7.8 sec.); Remaining (32 sec.)
170  | DES Log: 8 entries
```

Fig.11. The Result of the Simulation in a Log File

Similar to the methods of simulation in CloudSim system there were conducted three experiments with different conditions, which means that each time we use different number of levels of protection. After all the experiments in OPNET IT platform integrated editor there were built and compared these graphs (Fig. 12): efficiency server connection conditions and without protection, the results of the network with the use of protection and without them, the simulation results regarding the end user.

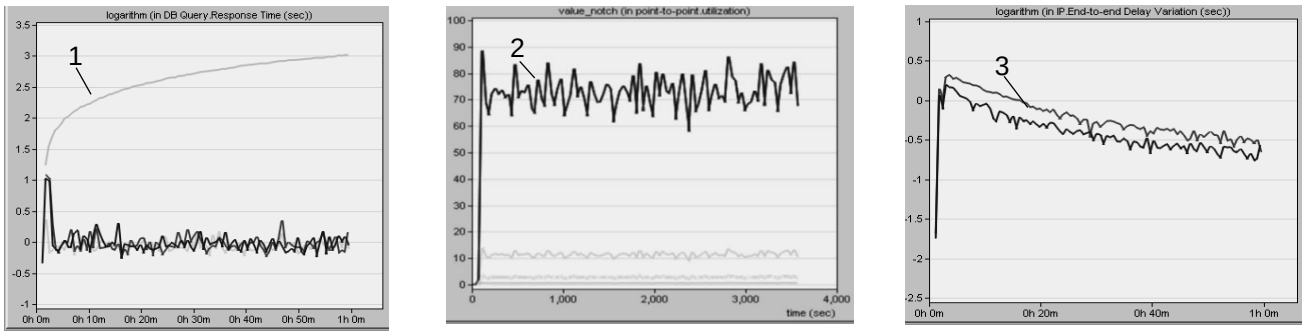
Fig. 12 (a) shows server performance curve with use of all security layers (1), and other curves represent performance without security layers connection. Therefore with the use of security means, performance and security of server are much higher.

Fig. 12 (b) shows network performance curve with use of all security layers (2), and other curves represent performance without security layers connection. Obviously, with the use of protection, efficiency and security of the network are much higher.

Fig. 12 (c) shows performance curve described data center work with end user with use of all security layers (3), and other curves represent performance without

security layers connection. Under the conditions of inclusion of protective efficiency of the data center with the end user is higher and is done through a secure communication.

After all relevant simulation results were compared to known models of data centers (Table 2) and determined that the model of a secure data center lacks identified in the gap analysis.



a *b* *c*

Fig.12. The Graphics Performance Of:Data Center Server (a), Network Data Center (b) Data Center With End User (c)

B. Experiment 2

The aim of the experiment: check the classification effectiveness of chosen anomaly detection method.

Input/output experiment data: input data – 10% of the KDDCup99 dataset, output data – assorted data (normal or abnormal). KDDCup99 – dataset used for The Third International Knowledge Discovery and Data Mining Tools Competition.

This database contains a standard set of data to be audited, which includes a wide variety of intrusions: DoS-attacks (denial of service); U2R-attacks (unauthorized access to local superuser privileges); R2L-attacks (unauthorized access from a remote machine); Probing-attacks (port scanning).

Steps of the experiment:

1. Loading the input data into the environment.
2. Choosing the classification algorithm, in that case – J48 – Java-implementation of decision tree algorithm.
3. Building the decision tree model.

In order to ensure, that chosen method is precise, the cross-validation mode was used with splitting dataset into 7 parts, 6 of which are the training dataset, and the remaining 1 – test dataset.

4. Revision of the experiment results: 4.1. Quantity and rate of correctly and wrongly determined data (fig.13). 4.2. Graphical look of the built tree (fig.14).

As a result of experiment, it is determined, that rate of correctly classified data is 99.96% that proves high accuracy and low false positive rate of the chosen algorithm. Also, graphical look of the built tree was considered.

Correctly Classified Instances	493820	99.9595 %
Incorrectly Classified Instances	200	0.0405 %
Kappa statistic	0.9993	
Mean absolute error	0	
Root mean squared error	0.0057	
Relative absolute error	0.0962 %	
Root relative squared error	3.5746 %	
Total Number of Instances	494020	

Fig.13. Rate of Correctly and Wrongly Determined Data

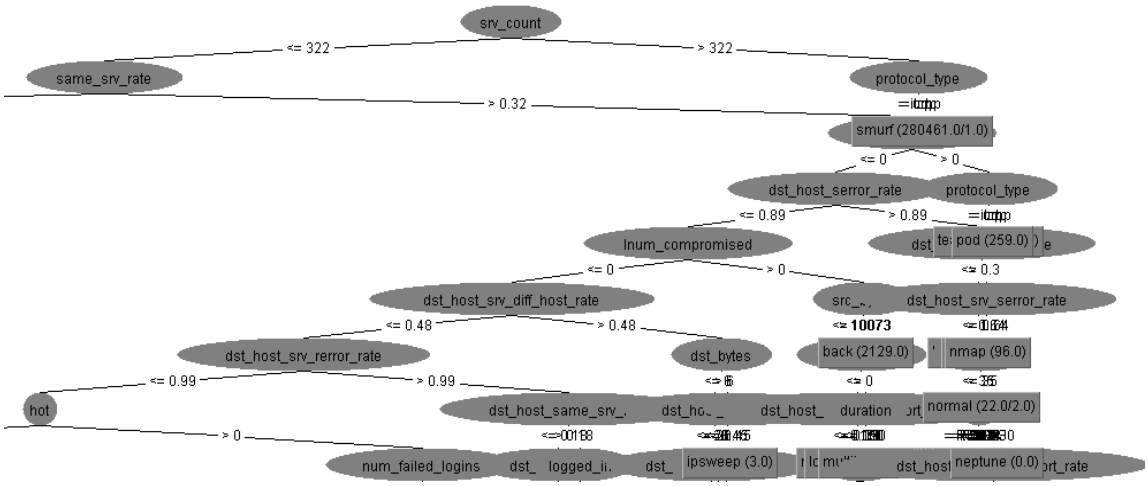


Fig.14. Graphic Expression of the Built Tree

V. Conclusion

In this paper there were selected and researched data center models based on Cloud Computing technology, revealed the problem of information security in almost perfect engineering and infrastructure solutions. Identified deficiencies have been remedied by developing of secure data center model based on Cloud Computing technology, which through the use of technology architecture, high-speed communications, unified computing structures and ensures security of the cloud data center and conduct appropriate simulation. The model can be used to build data centers in different areas. In addition, a model was developed for the detection of anomalies secure environment "cloud" computing based on the concept Big Data.

Also, there were done the analysis of modern methods of identifying anomalies and taking into account their shortcomings there was developed hybrid system anomaly detection that by using the method DecisionTree, signature module Snort, technology BigData (HDFS, YARN, MapReduce, Spark) and databases KDDCup99 can detect anomalies in traffic secure environment "cloud" computing; experimentally investigated anomaly detection module in the application Weka, which proved highly accurate algorithm. The practical value is the ability to integrate the developed system anomaly detection in the protected environment of "cloud" computing and increasing the percentage of detection through the use of signature module that can detect known attacks.

Acknowledgment

This scientific work was supported by RAMECS, CCNU16A02015 and Young Scientists Association of National Aviation University (Kyiv, Ukraine).

References

[1] A. Ghaffa, R. Soomro, "Big Data Analysis: Ap Spark Perspective", Global Journal of Computer Science and Technology: Software & Data Engineering, Vol., 15 Iss.1, 2015.

[2] Ah. Aljarray and Ab. Almadar, "Analysis and Detection of Fraud in International Calls Using Decision Tree", R&D Office, Libya-Misrata.

[3] V Chandola, A Banerjee and V Kumar, "Anomaly detection: A Survey", ACM computing surveys (CSUR), 2009.

[4] S. Bondarovets, O. Koval and S. Gnatyuk, "Anomaly Detection System For Mobile Carrier Based On Big Data Concept" in Information Technology And Security, vol. 4, no. 2, pp. 25-35, 2016.

[5] M. Boniface, "Platform-as-a-Service Architecture for Real-Time Quality of Service Management in Clouds", 5th International Conference on Internet and Web Applications and Services (ICIW (Barcelona, Spain: IEEE, 2010), pp 155–160.

[6] "Breaking down what's in your cloud SLA" [Online]. Available: http://searchcloudcomputing.techtarget.com/essentialguide/Breaking-down-whats-in-your-cloud-SLA.

[7] "Cloud Security". [Online]. Available: http://ru.thales-esecurity.com/solutions/by-business-issue/cloud-computing-security.

[8] "Data Center Architecture Overview". [Online]. Available: http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_1.html.

[9] "Data Center Design Models Overview". [Online]. Available: http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_1.html#wp1058588.

[10] "Data Center Tiers Explained", 2013. [Online]. Available: http://webcache.googleusercontent.com/search?q=cache:http://www.thedatacave.com/data-center-tiers-explained&gws_rd=cr&ei=VEvoVsSR KcL8swG78ZH4 BQ.

[11] M. Dodan, "Architected Cloud Solutions Revealed", Journal of Object Technology, vol. 9 (2).. pp. 27-36, 2010.

[12] "Explain: Tier 1 / Tier 2 / Tier 3 / Tier 4 Data Center".[Online]. Available: http://www.cybercitibiz/faq/data-center-standard-overview/.

[13] "Hybrid cloud: is it right for your business?", 2014. [Online]. Available: http://www.techradar.com/news/internet/cloud-services/hybrid-cloud-is-it-right-for-your-business—1261343.

[14] L. Kalinichenko A, I. Shanin and Taraban I "Methods for Anomaly Detection: a Survey", 16th Russian Conference on Digital Libraries RCDL Proceedings, 2014, pp. 20-25.

[15] O. Koval, S. Bondarovets and S. Gnatyuk: Secured data center model based on Cloud Computing technology", Ukrainian Information Security Research Journal, vol. 18, no. 2, pp. 133-143, 2016.

[16] "Models of cloud technologies", 2012. [Online]. Available: http://wiki.vspu.ru/workroom/adb91/index.

[17] "NIST Cloud Computing Reference Architecture", [Online]. Available http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/ReferenceArchitectureTaxonomy/NIST_CC_Reference_Architecture_v1_March_30_2011.pdf.

[18] "Security as a headache of cloud computing", [Online]. Available: http://www.cnews.ru/reviews/free/saas/articles/articles12.shtml

[19] "Security Recommendations for Cloud Computing Providers. White Paper", Federal Office for Information Security, GmbH.: Druckpartner Moser Druck, 2011.

[20] "Self-Run Private Cloud", [Online]. Available: http://www.gov connection.com/IPA/PM/Info/Cloud-Computing/Self-Run-Private-Cloud.htm.

[21] S. Sagiroglu and D. Sinanc, "Big Data: A Review" IEEE, 2013.

[22] "Snort Manual" [Online]. Available: http://manual-snort-org.s3-website-us-east-1.amazonaws.com/.

[23] "The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology", 2011. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

[24] "Threats of Cloud Computing and methods of their protection", 2015. [[Online]. Available: https://habrahabr.ru/post/183168/

[25] "Understanding the Different Types of Data Center Facilities" [Online]. Available: http://www.cyrusone.com/blog/under standing-the-different-types-of-data-center-facilities/.

[26] "What is cloud computing?" [Online]. Available: http://www.ibm.com/cloud-computing/what-is-cloud-computing.html.

[27] "What type of data center do you need? [Online]. Available:http://www.compassdatacenters.com/type-data-center-need/.
[28] "Whitepaper Cloud Computing Use Cases Version 3.0, produced by the Cloud Computing Use Case Discussion Group" [Online]. Available: http://opencloudmanifesto.org/cloud_computing_use_cases_whitepaper-3_0.pdf.
[29] "4 types of data centers", 2012. [Online]. Available: https://gigaom.com/2012/10/15/4-types-of-data-centers/.

Authors' Profiles

Zhengbing Hu PhD, Associate Professor of School of Educational Information Technology, Central China Normal University, M.Sc. (2002), Ph.D. (2006) from the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Postdoc (2008), Huazhong University of Science and Technology, China. Honorary Associate Researcher (2012), Hong Kong University, Hong Kong. Major research interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.

Sergiy Gnatyuk PhD, Associate Professor. In 2007 he received MSc degree in information security from National Aviation University (NAU, Kyiv, Ukraine). He received PhD in Eng degree in information security (quantum cryptography) from NAU in 2011. He is currently working at NAU in Academic Department of IT-Security. IEEE Member, Scientific Adviser of Engineering Academy of Ukraine, Executive Secretary of Ukrainian Scientific Journal of Information Security, Chairman in Young Scientist Association of NAU. Research interests: Cryptography, Quantum Key Distribution, Network & Internet Security, Information Security Incident Management, Cybersecurity & CIIP.

Oksana Koval Master's Degree Student. As a result of the Degree Thesis defense "Secured data center model based on Cloud Computing technology" in 2016 she received Bachelor's Degree in Information Security Management from NAU. Research interests: Information Security, Data Analysis, Cloud Computing, Cybersecurity, Information Security Management Systems.

Viktor Gnatyuk PhD Student (2012-2015), Assistant Teacher (from 2013). In 2012 he received MSc degree in Economic Cybernetic from Khmelnitsky National University (Khmelnitsky, Ukraine). He is currently working at NAU in Telecommunication Systems Academic Department. Research interests: Computer Network & Internet Security, Information Security Incident Management.

Serhii Bondarovets Master's Degree Student. As a result of the Degree Thesis defense "Anomaly detection system for mobile carrier based on Big Data concept" in 2016 he received Bachelor's Degree in Information Security Management from NAU. Research interests: Data Analysis, Cloud Computing, Cryptography and Cryptoanalysis, Public Key Infrastructure, Cybersecurity.

How to cite this paper: Zhengbing Hu, Sergiy Gnatyuk, Oksana Koval, Viktor Gnatyuk, Serhii Bondarovets,"Anomaly Detection System in Secure Cloud Computing Environment", International Journal of Computer Network and Information Security(IJCNIS), Vol.9, No.4, pp. 10-21, 2017.DOI: 10.5815/ijcnis.2017.04.02

