Development of Network Security Models in the Software-defined Infrastructure of Virtual Data Center

Irina P. Bolodurina [#1], Denis I. Parfenov [*2]
[#] Department of Applied mathematics; [*]Faculty of Distance Learning Technologies
Orenburg State University
Orenburg, Russia
[1]prmat@mail.osu.ru, [2]fdot_it@mail.osu.ru

***Abstract*— The novelty of the research is the combination of technologies of software-defined networks and virtualization of network functions for optimization network organization in virtual data centers. We was prepared approaches for solving tasks for describing firewall rules and the rules for selecting nodes of the physical infrastructure for the placement of security networks elements with using neural network and Data Mining methods. Our solutions will provide the required level of security and maintain a specified quality of service for applications and services of multi-cloud platforms. This approach was combined with efficient means for processing Big Data flows. It allows predicting infrastructure behavior over time and warning of possible failures, security threats and cyber attacks. The uniqueness our technology is that resource management that provides security and quality of service is carried out autonomously, using the principles of self-organization, without interference from administrators and third parties.**

Keywords—multi-cloud platforms, software-defined networks, software-defined infrastructure, virtual data center, virtual network functions

I. INTRODUCTION

Currently, significant amounts of convergent traffic are circulating within the virtual data center. The creation of effective mechanisms for managing and protecting such traffic flows in a data center network requires an understanding of its structure, as well as the definition of the load on its computing and network nodes when it is serviced. To solve this problem, within the framework of the research, the task was solved to develop an algorithm for firewalling and to ensure the quality of servicing of data flows in a network environment of a virtual data center. The solution proposed by us is based on an approach that allows aggregating all transmitted flows in a single analytical system for collecting SIEM data. The basic element of the proposed approach is a distributed network consisting of virtualized modules that perform processing and analysis of traffic passing through them. Due to decentralization, the proposed solution allows for unlimited scalability. All collected and analyzed information is submitted with aggregated and compressed form to a single network management center, where they are converted into the corresponding controller rules according to the OpenFlow standard.

The presented research is aimed at solving problems related to improving the efficiency of quality of service provision, as well as the security of the software-defined infrastructure of multi-cloud platforms located in virtual data centers.

II. PROBLEM FORMULATION

The presented research developed the architecture and the algorithmic solutions for an autonomous self-organizing security system, quality of service, detection of flows and protection from cyber attacks in the software-defined infrastructure of a multi-cloud platform located in a virtual data center. The ensemble of models allows us to describe its structure, as well as the principles of its operation. The originality of the developed solutions consists in using the technology of containerization of basic services and services, as well as the use of software-defined networks and virtualization network functions. The mathematical models presented in the article allow us to describe the self-organization of data flow control and firewall rules and the choice of physical infrastructure nodes for the placement of security elements to ensure quality of service and network security in the virtual data center. This approach allows you to fine tune the rules for filtering and classifying packets.

One of the most important tasks for increasing the efficiency of the network is to analyze the performance of all the components, which support the work of the data center. A significant difference of the virtual data center is the use of virtual networks for routing data flows of cloud applications and services over the existing physical infrastructure. Virtualization at the network level allows us to run simultaneously many different types of applications with different performance requirements for compute nodes, network objects, and QoS [11, 12]. Therefore, we should be aware of the flows circulating in the network at present time to provide the information about the quality of service at the network of the virtual data center. Most modern methods of

The research work was funded by Russian Foundation for Basic Research, according to the research projects No. 16-37-60086 mol_a_dk, 16-07-01004 and the President of the Russian Federation within the grant for state support of young Russian scientists (MK-1624.2017.9).

 978-1-5386-4340-2/18/$31.00 ©2018 IEEE

traffic classification define the flow as a group of packets, which have the same destination IP address and use the same transport protocol and port numbers. This definition allows considering bi-directional flows. Therefore, the packets that transfer requests from the user and answers from compute nodes are a part of the same flow in a network of the virtual data center. The approach to classification based on selected characteristics is a fairly simple task; however, the accuracy of such solutions for modern cloud applications and services are not quite high. In terms of architecture, modern cloud applications are comprehensive objects consisting of many clustered and distributed services. In turn, each cloud service used by the cloud application has its own set of requirements for QoS and dependencies. In the virtual data center, the procedure of user request processing can be represented as a multi-phase queuing system, where data flows are described by different laws of distribution. Thus, the problem of analysis and classification of the traffic flows of cloud applications and services in the virtual data center becomes non-trivial. Another feature of using a software-defined network is the use of dynamic port numbers for routing flows for the same type of cloud application. This feature makes it difficult to classify the traffic of applications by this attribute. The existing solutions based on deep packet inspection (DPI) work slowly and require a lot of processing power.

In this investigation, we propose an approach for accelerating the learning process and improving the accuracy for traffic flow classification of cloud applications and services in a network environment of the virtual data center at the initial stage of data analysis.

III. Model of Self-Organization of Data Flow Control

The model of self-organization of data flow control is created on the basis of the decomposition method and approaches used in queuing theory. To determine the management objects at the input of each virtual network node, sparse traffic outflows from other nodes aggregate. The solution of systems of linear algebraic equations allows determining the intensities and average values of time intervals between neighboring packets for each flow. The flow of packets coming from an external source to the network is described by the law of distribution of time intervals between neighboring packets. This means decomposition of the network into separate nodes, which allows you to identify all the main traffic flow indicators and provide the required quality of service within the virtual data center network.

Among the most important quality-of-service (Q) targets are usually the following set of characteristics:

$$Q = \{AvgA,\ MinU,\ AvgRT,\ MaxU,\ AvgB,\ IR\}, \quad (1)$$

where AvgA – average availability, expressed as the average number of failures during the service provision period; MinU – minimum availability for each user; AvgRT – average service response time; MaxU – maximum response time for each user; AvgB – average throughput; IR – methodology for calculating metrics and frequency of collection of reports.

To improve the quality of services and the security of multi-cloud platforms, we need to detail the main object, which should give a prediction about the architecture of the network of the virtual data center. The network of the virtual data center can be described as

$$VDCnetwork = (V, e, u, FE), \quad (2)$$

where V – vertices of the graph (network nodes), e – arc graphs (network connections), $FE : E(G) \to R^+$ - flow of requests in a virtual data center network; $u : E(G) \to R^+$ - the cost of implementing a network connection in a data center network.

To maintain the required quality of service (QoS) and a given level of uninterrupted operation, we will write the balance equations:

$$ex_f(u,v) := \sum_{e \in \delta^-(U,V)} FE(e) - \sum_{e \in \delta^+(U,V)} FE(e), \forall e = (u,v) \in E(G) \quad (3)$$

The presented description is necessary for transition to modeling at the level of flows of transmitted and analyzed data. Each record of the flow FE_{kij}= $FE_{ki}(t)$ is dynamic and changes at times t. Each thread has a set of characteristics that uniquely identify it. In the traffic management system, the flow can be represent as form

$$\mathrm{FE}_{\mathrm{kij}} = (\mathrm{Match}_{\mathrm{kij}}, \mathrm{Act}_{\mathrm{kij}}, \mathrm{TimeOut}_{\mathrm{kij}}, \mathrm{Flow}_{\mathrm{kij}}, \mathrm{Cnt}_{\mathrm{kij}}(t)), \quad (4)$$

where $Match_{kij}$ this is a set of fields to check for matches with the headers of the package; Act_{kij} – A set of actions performed on the package, if its headers match $Match_{kij}$; $TimeOut_{kij}$ – time of fixing the flow in the system; $Flow_{kij}$ – the thread to which this OpenFlow rule applies; Counters – statistical counters OpenFlow.

In this case, for more efficient traffic analysis, as a rule, a signature approach is used in networks. Within the framework of this research, this approach is supplemented by methods of data mining, which allow reducing the set of characteristics, which significantly speeds up the processing of information. This is important when analyzing Big data flows that occur in networks of virtual data centers because of multiple intersections of communication channels. The flow analysis scheme at the point of connection to the network, on the node of the data center, can be formalized as follows:

$$\mathrm{Analyzer}_{\mathrm{ki}} = (\mathrm{Node}_{\mathrm{ki}}, \mathrm{CuFlows}_{\mathrm{ki}}(t), \mathrm{SusFlows}_{\mathrm{ki}}(t)), \quad (5)$$

where $Node_{ki}$ – the network node on which the signature traffic analyzer works, $CuFlows_{ki}(t)$={$CurrentFlows_{ki}$} and $SuFlows_{ki}(t)$ = {$SuspiciousFlow_{ki}$} – respectively, the set of current and the set of suspicious flows detected by the analyzer at time t.

For the model presented, firewall rules have been developed separately for the L2 and L3-L4 levels of the OSI model. All rules have two parts - the headers that identify the

packages, and the action. The headers for L2-level rules include the source port of the packet, the source switch, the MAC addresses of the sender and receiver of the packet, the type of protocol, etc. The headings for the L3-L4 layer rules contain the IP addresses and ports of the source and destination of the packet, the type of the encapsulated protocol, etc. Possible actions - the removal or resolution of the package.

Such rules can be combined into chains if necessary. Each chain is an ordered list of rules that has an identifier. The entire chain of rules is checked in order of priority, until there is a rule suitable for the parsed packet. In this case, the action specified in this rule is executed, and the subsequent execution of the chain is terminated. In order to transfer the rule to another chain, the corresponding action with its identifier can be specified. The originality of this model lies in the fact that within the rules all OpenFlow headers are supported up to version 1.5, inclusive.

IV. The Algorithm of Adaptive Firewall and Ensuring the QoS

From an algorithmic point of view, we can represent classification or clustering of a traffic flow as the function $f: X \to C$, which puts the label $c_j \in C$ in correspondence with each object $x_i \in X$. The set of C is defined in advance. In the task of clustering, neither the set of C nor its dimensionalities is determined. In this research, we have used the classification and clustering of traffic flows of the cloud applications and services to improve the efficiency of QoS inside the virtual data center.

A classification traffic flows of cloud applications and services located in the virtual data center can be divided into the following elements: classification, clustering and identification of association rules. Let's consider these elements of the model separately.

For the effective classification of traffic flows in the virtual data center, we need to determine the set of all applications. Suppose that we have set of cloud applications and services defined as $X = \{x_1, \ldots, x_n\}$. Each cloud application and service is characterized by a set of attributes $X_j = \{a_1, \ldots, a_m, y\}$, where a_i is the observed attributes, whose values represent characteristics of the traffic of cloud application or service; y is the target attribute that identifies the class of a cloud application or service. Each attribute a_i takes a value from some set $A_i = \{a_{i,1}, a_{i,2}, \ldots\}$, which describes valid values of characteristics of the attributes in the subject area under study. In the framework of solving the tasks of traffic classification, suppose a limited number of classes applications $y \in C$ circulate in the network of the virtual data center, where $C = \{c_1, \ldots, c_k\}$. With regard to our task, we will explore the traffic flows of the cloud applications and services according to the communication scheme of their interaction with the network objects within the virtual data center.

The algorithm of the adaptive firewall for software-defined infrastructure with multi cloud platform, taking into account the context (state) of the data transmission. The main idea of this algorithm is the following: if the packet arrives at the OpenFlow controller, and then, if inside IP protocol is encapsulated, then for L3-L4. The algorithm supports sessions (status monitoring) for TCP and UDP protocols. The beginning of the session for the TCP protocol is monitored by the first step of the three-step handshake (the packet with the SYN flag), and the end by the timeout in 300 seconds or when receiving packets with FIN and RST (in this case, the timeout is adjusted to 60 seconds). For the UDP protocol, the beginning is for the first packet, the end is only for a timeout of 300 seconds.

Each OpenFlow rule that is implemented in the OpenFlow is disabled in the OpenFlow, which allows optimizing the limited sizes of the thread tables OpenFlow switches.

The main advantage of the algorithm is the closest to the packet receiver, and then on the switch closest to the packet sender. In general, the algorithm can lock on any OpenFlow switch, any analyzing inter-node traffic. However, with this approach, the performance of the corporate network may decrease, therefore, within the framework of this research, it is proposed to use the developed firewall. The generalized algorithm for adaptive firewall and ensuring the QoS in virtual data center can be represented as the following sequence of steps:

Step 1. To identify the traffic flows, we use the data received from the controller of a software-defined network, which controls the placement of cloud applications and services in the virtual data center.

Step 2. Basing on the obtained data, we formed a graph in compliance with the communication schemes of interaction between cloud applications and services.

Step 3. The obtained schemes are overlaid on the current topology of the software-defined network to evaluate the network bandwidth, the usage of virtual channels and the analysis of primary transmitted thereon data basing on moments of the packets reception time distribution.

Step 4. The data obtained are ranked in descending order by the load of the communication channel and by the priority of the traffic flows of cloud applications. To adjust the routes, 20% of the most loaded channels are selected.

Step 5. For the selected virtual channels, the traffic flows of cloud applications are classified more thoroughly to identify the degree of channel usage by particular types of applications.

Step 6. For the applications identified in the previous step, we analyze the traffic, get the parameters of state from physical network devices and identify the most loaded objects.

Step 7. High loaded devices are excluded from the current route and the traffic is redistributed between less loaded nodes by using association rules for routing. These changes are also updated in the communication schemes of interaction between applications and services.

Step 8. The results of the analysis are used to apply QoS policies on the controller of a software-defined network in the virtual data center as well as in a retrospective analysis of data for error correction.

Thus, the main goal of the developed algorithm is to find the optimal solution and to maximize the performance of the physical network taking into account the existing flows of applications and services and their demands for delays in the work of the virtual data center.

V. Experimental results

To assess the effectiveness of the developed algorithm for optimizing the adaptive routing of data flow balancing in the applications and services in the virtual data center, we have conducted a pilot study. We have chosen the Openstack cloud as the basic platform. For comparison, we have applied the algorithms used in the OpenFlow version 1.4, for route control of the software-defined network in the experiment. For the experimental research, a prototype has been created, including basic nodes, as well as software modules for the developed algorithms that redistribute data flows and applications. To verify the developed algorithm of optimal routing and traffic balancing in case of dynamic changes channels in a software-defined network of the virtual data center, several experimental networks consisting of 25, 50, 100, 200, 300, and 400 objects have been deployed. All generated requests were played consequently on two pilot sites: the traditional routing technology (Platform 1, NW) and the technology of the software-defined networks (Platform 2, SDN). This restriction is caused by the need to compare the results to a traditional network infrastructure, which is not capable of dynamic reconfiguration. Two tests were carried out on site 2. In the first case, the model OpenFlow version 1.4 routing algorithms were in use, in the second case (Platform 3, NEW SDN), the developed routing optimization algorithm was applied. The experiment time was one hour, which corresponds to the most prolonged period of peak demand, recorded in a real traffic network of a heterogeneous cloud platform. We have chosen response time of applications and services that work in a cloud platform as a basic metrics to assess the efficiency of the proposed solutions. The results of the experiment are provided in Fig. 1.

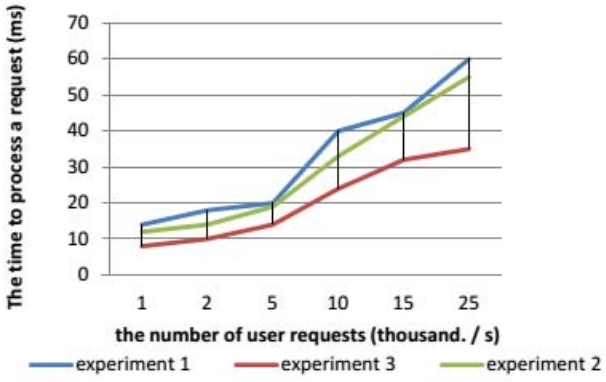

Fig. 1. Computational experiment result

VI. Discussion

In article [1], the problem of responding to an incident in the cloud is considered. An incident is an approach to eliminating and control the consequences of a security breach or a network attack. As a key criterion of the model authors select the minimization of the incident processing time is considered due to the introduction of security controllers and security domain in the cloud infrastructure for the analysis of network threats.

To counter network cyber attacks, technologies began to appear to provide continuous monitoring on any device connected to the network in order to increase fault tolerance by detecting and mitigating targeted threats. The authors of the research [2] describe the Gestalt security architecture, which is based on the principles of strong isolation, the policy of least privileges, the concept of providing information security (defense-in-depth), cryptographic authentication, encryption and self-healing. Remote monitoring is carried out through an organized workflow through a multitude of components connected by a specialized secure communication protocol that together provide secure and sustainable access.

Conventional firewalls are used to enforce network security policies on the boundaries within the network. In [3], the authors take advantage of the SDN to turn the network infrastructure into a virtual firewall, thereby improving network security. The virtual firewall as ACLSwitch is presented, which uses the OpenFlow protocol to filter network traffic between OpenFlow switches. The authors also introduce domain policies that allow the use of different filtering configurations for different network switches.

The optimal use of IDS and IPS systems is also presented in [4]. The authors investigated the open source snort system. The implementation, tuning, installation of the system and the problems arising during the research are studied.

Firewalls often have vulnerabilities that can be exploited by cybercriminals. The publication [5] is devoted to the investigation of some possible fingerprinting methods (fingerprint, identification) of the firewall, which turned out to be quite accurate. The authors also studied DoF attacks (Denial of Firewalling) in which attackers use carefully processed traffic to overload the firewall. In [6] the implementation of a third-level firewall with a full-mesh topology with 1 controller and 6 switches is presented. Modification of the learning switch code for the POX controller is performed for the full-mesh topology inside the Mininet network emulator. In this case, the packet flow between hosts is controlled in accordance with the rules recorded in the learning switch through the OpenFlow controller.

In the context of the joint use of SDN and NFV, the NetFATE architecture is presented in [7], a platform designed for placing virtual network functions on the network boundary. This architecture is based on free open-source software on the provider's nodes and client equipment, which leads to a simpler deployment of functions and lower management costs.

An important part of any network architecture is application identification, which contains information about the activity of network applications, and their use of bandwidth. The article [8] focuses on the promotion of open source technology to identify applications. This technology is seen as the future of firewalls, where the administrator can view more detailed

information about network traffic compared to current approaches.

The issues of achieving elasticity for network firewalls are discussed in detail in [9]. Elasticity here means the ability to adapt to network load changes by releasing and allocating resources in an autonomous manner. The elasticity of cloud firewalls is aimed at satisfying a consistent performance evaluation using the minimum number of instances of the firewall. The author's contribution of the publication is to determine the number of instances that must be dynamically adjusted in accordance with the load of incoming traffic and the base of firewall rules.

Due to software processing of network functions, the performance of VNF significantly decreases depending on the types of VNF and the configuration of VNF applications. In the publication [10], the authors pay special attention to the analysis of the virtual firewall as a representative of VNF. The paper proposes a method for estimating the latent load of a virtual firewall using rules in the ACL and the amount of traffic for each rule.

VII. Conclusion

The developed methods and algorithms allow to increase the level of research in the field of information security and quality of service in multi-platform platforms in the virtual data centers. The application of the proposed approaches based on the joint use of virtualization network functions and software-defined networks allows for more efficient planning of data flows and providing the required quality of service and a given level of security in the data center network. Developed models and algorithms have a high innovative potential and can serve as a basis for developing an industrial system for detecting threats and protecting against cyber attacks in virtual data centers.

REFERENCES

[1] *Alexander Adamov, Anders Carlsson* Cloud incident response model // Proceedings of 2016 IEEE East-West Design & Test Symposium (EWDTS) – 2016 – P. 1-3.

[2] *Michael Atighetchi, Aaron Adler* A Framework for Resilient Remote Monitoring // Proceedings of 2014 7th International Symposium on Resilient Control Systems (ISRCS) – 2014 – P. 1-8.

[3] *Jarrod N. Bakker, Ian Welch, Winston K.G. Seah* Network-wide Virtual Firewall using SDN/OpenFlow // Proceedings of 2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN) – 2016 – P. 1-7.

[4] *Mhair Kashif, Zahoor-ul-haq* An Optimal Use of Intrusion Detection and Prevention System (IDPS) Proceedings of 2015 European Intelligence and Security Informatics Conference – 2015 – P. 190.

[5] *Alex X. Liu, Amir R. Khakpour, Joshua W. Hulst, Zihui Ge, Dan Pei, Jia* Wang Firewall Fingerprinting and Denial of Firewalling Attacks // IEEE Transactions on Information Forensics and Security Journal – 2017 – V. 12 – No. 7 – P. 1699 – 1712.

[6] *Avinash Kumar, N. K. Srinath* Implementing a firewall functionality for mesh networks using SDN controller // Proceedings of 2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions – 2016 – P. 168-173.

[7] *A. Lombardo, A. Manzalini, G. Schembra, G. Faraci, C. Rametta, V. Riccobene* An Open Framework to Enable NetFATE (Network Functions At The Edge) // Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft) – 2015 – P. 1-6.

[8] *Nainesh V. Patel, Dr. Narendra M. Patel, Costas Kleopa* OpenAppID – Application Identification Framework // Proceedings of 2016 Online International Conference on Green Engineering and Technologies (IC-GET) – 2016 – P. 1-5.

[9] *Khaled Salah, Prasad Calyam, Raouf Boutaba* Analytical Model for Elastic Scaling of Cloud-based Firewalls // IEEE Transactions on Network and Service Management Journal – 2016 – V. 14 – No. 1 – P. 136-146.

[10] *Dai Suzuki, Satoshi Imai, Toru Katagiri* A New Index of Hidden Workload for Firewall Rule Processing on Virtual Machine // Proceedings of 2017 International Conference on Computing, Networking and Communications (ICNC): Communications QoS and System Modeling – 2017 – P. 632-637.

[11] Parfenov D., Bolodurina I. “Methods and algorithms optimization of adaptive traffic control in the virtual data center” 2017 International Siberian Conference on Control and Communications, SIBCON 2017 - Proceedings 29-30 June 2017, Astana, Kazakhstan. 2017. - P. 1-6.

[12] Bolodurina I. P., Parfenov D. I. Development and Research of Modelsof Organization Distributed Cloud Computing Based on the Softwaredefined Infrastructure // Procedia Computer Science — Vol. 103 —P. 569–576.