Approaches for Optimization Using Virtual Network Functions in Infrastructure of Virtual Data Center

Irina Bolodurina[1] and Denis Parfenov[2]

[1] Orenburg State University, Department of Applied Mathematics, Orenburg, Russia
[2] Orenburg State University, Faculty of Distance Learning Technologies, Orenburg, Russia
prmat@mail.osu.ru

Abstract. The study proposed a classification and identification model of virtual network functions based on the statistical properties of the flow and defined a systematic approach to the selection of the optimal set of attributes of the traffic flow. The approach applied in our investigation for placement of virtual network functions allows to optimizing traffic flows in virtual data center. It includes algorithmic solutions based on neural networks allowing to identify network functions.

Keywords: multi cloud platforms · network function virtualization · software-defined networks · virtual data center

1 Introduction

Today, considerable volumes of data circulate in modern telecommunication networks. The data centers are the nodes of aggregation and flow processing. The modern paradigm of the network calculation environment demands the introduction of adaptive and flexible solutions. New solutions need effective control of traffic in networks and, at the same time, do not demand cardinal changes of the existing infrastructure of a data center. Traditional data centers use the concept of resources virtualization for various infrastructure facilities (network, computing nodes, systems of storage and others) to achieve the goal. The application of a complex approach to virtualization is reflected in the architecture of virtual data centers. This architecture was used for the placement of multi cloud platforms. Multi cloud platforms use hybrid methods of virtualization based on software-defined components. It enables to increase the efficiency of computing resources use and, thus, to reduce the economic cost of maintaining the infrastructure of traditional data centers. However, the conception of resources virtualization is not quite effective. It allows abstracting the processed and transmitted data flows from physical devices. [1]. But, nowadays, the problem of the effective placement of key components of the virtual network environment in a multi cloud platform is not solved.

One of the approaches applied in virtual data centers apart from the virtualization of traditional objects of network infrastructure, is the use of software

realization instead of traditional hardware solutions, such as firewall, load balancer, NAT, routers and others [3]. In practice, such solutions are based on the technology of network function virtualization (NFV). The NFV technology provides more flexible deployment and enables to control the virtual objects of a multi cloud platform, which perform the roles of hardware network devices, more effectively. As a rule, the NFV technology is applied together with the software-defined network and enables to exercise adaptive traffic control. However, the technology of network function virtualization has a number of disadvantages. The main problem is the lack of effective methods of planning for placing virtual objects in physical computing nodes. The review of research shows that existing solutions for placing the NFV in the infrastructure of data center use the approaches based on virtual machines or containers [2]. The existing solutions do not deal with resource intensity of each virtual network function and its functional purpose for multi cloud infrastructure of a virtual data center. We have developed the approach that allows us to cluster the existing virtual and physical objects of infrastructure and, then, to place virtual network functions. The main idea of our solution is to estimate the consumption of resources by each element of the network. Besides, we will use the hybrid method of virtualization based on the simultaneous use of virtual machines and containers to create a flexible solution. It will enable to optimize the placement of the technology of network function virtualization in the infrastructure of a virtual data center.

Our approach is relevant, since it represents the combination of two modern innovative technologies in the field of the organization of network functioning and virtualization of its components for resource and data flow control in the software-defined networks based on the technology of network function virtualization. The goal of our investigation is to improve the quality of service for applications and services of the multi cloud platforms placed in a virtual data center. Besides, we use the methods of intellectual data analysis to process information about the state and load of key objects as well as the flows between network devices received from the systems of computing nodes monitoring in the software-defined infrastructure of the multi cloud platform. It enables to receive the consolidated assessment of the quality of service and to predict uninterrupted operation and operability of the software-defined infrastructure of a multi cloud platform and the entire virtual data center.

In the following sections we'll describe the approaches developed by us directed to a solution of the problem of optimization placement of virtual network functions in the multi cloud environment of virtual data center. Section 2 gives information on a condition of researches today and the existing approaches in work with network functions. In Section 3 we'll describe the methods and approaches applied in our decision, and we'll also describe the main stages of its realization. The neuronetwork model which is a basis for formation of cards of placement of network functions in the multi cloud environment of virtual data center is presented in section 4. In section 5 the algorithmic solutions and experimental explorations which in practice describing an optimization task placement of network functions. Conclusions is presented in section 6.

2 Related Work

The variety of physical network devices of various vendors increases both capital expenditure and operational costs for the maintenance of data centers. The technology of network function virtualization allows solving this problem by the realization of network functions as software. The application of the NFV implies the use of the technology of network objects virtualization, which function as software and particular computing processes, or as complex infrastructures of cloud computing instead of hardware solutions.

The group of scientists headed by Min Chen [3] analyzes the architecture and mechanisms of the interaction between the technology of network function virtualization and the software-defined networks. As noted in the research, if the number of the users who launch applications in the multi cloud network environment increases, there is a competition for resources. Besides, each user request is described by the relevant requirements to network environment from the viewpoint of productivity, safety, and the effective control of objects in virtual infrastructure.

Scientists from Arizona State University [4] have studied a multi cloud system. They offer an approach to the creation of network architecture based on the NFV technology as alternatives for traditional hardware network devices. However, this research does not solve the main problem concerning the methods of NFV placement on computing nodes.

Apart from a problem of the effective placement of network functions, the NFV technology has a number of disadvantages associated with the organization of the coordinated control of the entire network infrastructure of a virtual data center [5]. To solve this problem, scientists from University of Wisconsin-Madison have proposed an approach, which is a framework of the OpenNF. This approach provides the effective coordinated control of both the internal state of network functions and the state of data transmission network. However, this decision does not solve another important issue associated with the overall performance and load on the controller and objects of network functions.

The scientists from Nation Chung Cheng University have investigated the reduction of load on the controller to ensure the work of the NFV technology [6]. As a rule, in case of the software-defined network use, the controller classifies the traffic received from the ports on network nodes to define a path to network objects, which play a role of the virtual network functions. This process generates large volumes of traffic in the plane of control. The authors offer the expansion for the architecture based on the software-defined network to reduce the load created by network traffic on the controller due to the use of the NFV technology. The solution represents two-layer classification of traffic based on the OpenFlow protocol. Network events are analyzed in the plane of data instead of the plane of control.

The group of scientists headed by Hassan Jameel Asghar has developed a scalable system to work with the technology of network function virtualization in the multi cloud environment [7]. The authors have offered an abstract model of the standard network function distributed between several cloud platforms.

The developed model is used as a basis for the SplitBox system. This system enables to increase the speed of package processing considerably in comparison with similar hardware solutions. An essential disadvantage of the concept is the resource intensity of this system. The overhead costs of the deployment and work of the Splitbox network functions demand the same quantity of computing resources as in case of hardware network devices. This disadvantage neutralizes the available advantages of the approach, because the problem of the effective placement of network functions in the multi cloud infrastructure of the virtual data center remains unsolved.

Thus, it is established that the technology of network function virtualization and the existing cloud solutions on based on the software-defined infrastructure of the virtual data center has a number of advantages, which enable to improve the quality of service in data transmission networks. However, today, there are adequate and effective solutions, that would enable to control the placement of the NVF on physical and virtual computing nodes in the data center.

3 Methods and Approaches

Nowadays, neural networks are the most effective and high-speed method for forecasting, parameter identification, clustering and classifications in various fields of knowledge. Today, we see many successful examples of the application of a neural network approach for the creation of intellectual information systems. Besides, the advantage of the neural networks use is the possibility of adaptive self-training with the use of additional methods of approaches. We have used an iterative approach based on a group of methods associated with the optimization of placement of virtual network functions on the objects of the software-defined infrastructure of the virtual data center.

First of all we will present all network objects of the multi cloud platform placed in virtual data center as a communication graph. The graph is based on the topology the physical network switching. Each network object is the graph vertex. It can be described by a basic set of parameters, which characterize each element of the network and influence productivity. We have chosen the following characteristics as parameters: volume of memory, volume of disk space, the frequency and quantity of kernels, etc. Further, we will use these characteristics as the input parameters acting as the training set for a neural network during the study of data.

A multi cloud platform supports the placement of various applications and services. Therefore, to identify the flows of traffic passing through infrastructure facilities of the data center is an important task for the placement of virtual network functions. In this research, we have used the method based on the analysis of the known network ports for popular applications to obtain this information. This method enables to make an integrated classification of traffic flows; however, since there are non-standard network solutions applied in service-oriented applications, there obtained data will not be enough for the effective control of traffic flows. For a deeper analysis of traffic flows as a data source, we

offer to use the method of decoding the protocols of communication based on the analysis of contents of the transferred packages. However, since this approach has rather high resource intensity, it will be used only at a low level of the analysis, for more exact identification of traffic flows of similar applications. The third method uses the sample approach based on the specific signatures located in protocol heading for the identification of the application. The fourth method is based on machine training. This method uses the accumulated data obtained by the above-mentioned methods and applies to them the algorithms of machine training to identify the applications based on characteristic packages and the saved-up statistics of data flows. The advantage of this approach is that algorithms can be trained in real time that will allow reconfiguring software-defined infrastructure of virtual data center on the fly. Use of the proposed solution based on an integrated approach to collection of data on the traffic flows circulating in a multi cloud platform. It will allow optimizing placement of network functions on computing nodes of virtual data center.

To achieve the goals of the research, we have created a neural network system to predict the placement of network function virtualization in the multi cloud environment of the virtual data center. This implies the consecutive implementation of a number of algorithmic and software solutions. First of all, the module of data collection, which enables to receive the sets of primary data about the state of the network infrastructure of the virtual data center, is implemented for a neural network system. The obtained information is necessary for neural network training and testing. The next stage is to use the obtained data to define the optimum scheme for the placement of network function virtualization and to carry out experimental approbation on a-priori known samples and the obtained results. This will allow us to correct the sets of input data and to improve the quality of obtained results at the neural network exit. The final stage is to test the system using the examples, which are not included in the training sample. This will enable to ensure the efficiency of the obtained results.

4 Model of Clustering and Identification Virtual Network Functions

We have chosen Kokhonen's network as a neural structure for modeling, since it is the most efficient in the clustering and classification of objects. Another important factor is the visualization of results; it enables to improve the understanding of the structure and character of data at early stages and to specify a neural network model further. Due to the peculiarities of network function virtualization, the support of classification in Kokhonen's network can be used to identify the uniform elements in network for further optimization of their placement. Kokhonen's network is trained by a method of consecutive approximations. Starting from the initial placement of objects selected randomly, the algorithm gradually improves it to supply the data clustering. Another advantage of Kokhonen's network is the opportunity to identify new clusters. The trained network detects clusters in the training data and refers all the data to

certain clusters. If the network meets a set of data, which differ from any known samples, it will independently reveal a new cluster of elements then. This feature is very relevant, since it allows entering new network functions into the architecture of virtual data center without the actual change of algorithms of their distribution on physical and virtual computing nodes.

The principle of creation of neural network system for optimization of placement of network function virtualization in a multi cloud environment of virtual data center is as follows. On the basis of the data obtained from systems of monitoring of virtual data center we have select a number of criteria. That possible to identify unambiguously the network function placed on computing nodes. Criteria are formulated so that the answer could be presented in the binary form that is 1 - "Yes" or 0 - "No". On the basis of the obtained data the table which moves on an entrance of neural network is formed. Also the vector of output values has a similar appearance. Its components also have a binary appearance.

We developed an algorithm of training of Kokhonen's network. Kokhonen's network consists of one layer of neurons. The number of entrances of each neuron is equal to n - it is total of the possible characteristics peculiar to network functions. The amount of neurons of m coincides with the required number of classes into which it is necessary to break, corresponds to the number of the unique network functions used in work of a multi cloud platform. The importance of each of entrances to neuron is characterized by the numerical size called by weight. Describe the training of Kokhonen's network by steps.

Step 1. Initialization of network.

Small casual values are appropriated to weight coefficients of network $w_{i,j}, i = \overline{1,n}, j = \overline{1,m}$.

Values a_0 -initial rate of training and D_0 - the maximum distance between weight vectors (W matrix columns) are set.

Step 2. Presentation of a new entrance signal to network of X.

Step 3. Calculations of distance from an entrance X to all neurons networks:

$$d_j = \sum_{n}^{i=1} \left(x_i - w_{i,j}^N\right)^2, j = \overline{1,m} \tag{1}$$

.

Step 4. The choice of neuron k of $1 \leq k \leq m$, with the smallest distance of d_k.

Step 5. Control of scales of neuron of k and all neurons which are from it at the distance which isn't surpassing d_N.

$$w_{i,j}^{N+1} = w_{i,j}^N + a_N \left(x_i - w_{i,j}^N\right) \tag{2}$$

.

Step 6. Reduction of values a_N, d_N.

Step 7. Steps 2-6 repeat until then, weight won't cease to change (or still total change of all scales will become very little).

After training classification is executed by giving on an entrance of network a vector to be examined, calculation the distance from it to each neuron with the

subsequent choice of neuron with the smallest distance as indicator of the correct classification. For training of a neuronet we have used the data obtained from system of monitoring of virtual data center of the Orenburg state university. For experimental research we have selected key network functions which using in all data centers to build standard network solutions. This allowed us to test the functioning of neural networks and to determine the correct identification of virtual network functions in the real system. In table 1 examples of experimental recognition of virtual network functions in a software-defined infrastructure on virtual data center are given.

Table 1. Result of experimental recognition of virtual network functions

Virtual network function	The number of instances	The number of right recognized
Router	20	19 (98%)
NAT	15	13 (94%)
Firewall	18	17 (93%)

On the basis of the obtained data it is possible to make a conclusion that for the constructed neurosystem presented certain difficulties of recognition of a number of network functions in view of their insignificant differences in the chosen parameters. This defect can be eliminated by introduction of additional criteria to initial model of neural network. Thus, application of the developed neural network system gives the chance to correctly identify virtual network functions in 93-98% that, promotes increase in efficiency the solution of a problem of optimization for their placement in infrastructure of a multi cloud platform.

5 Algorithm of Optimization of Distribution of Virtual Network Functions in Infrastructure of Virtual Data Center

The presented model of identification of virtual network functions allows optimizing their placement in virtual data center. We will perform optimization of placement the network functions which found by using Kokhonen's network in virtual data center by the following criteria: the current load which created on computing nodes; resource intensity of network function; quantity of the flows passing through computing nodes. The main objective of placement of virtual network functions is the choice of optimum quantity of the nodes which realizing required functionality as software solution. Thus, the problem of planning of resources takes place. At the organization of dynamic topology in virtual data center planning is particularly relevant. The created load of computing nodes can change over a wide range for rather short intervals of time and depends on the chosen type of placing specific network functions. The algorithms of the monitoring of infrastructure of virtual data center, placement and start of network

functions are developed for the solution of an optimizing task. In comparison with the available analogs the algorithm uses the heuristic analysis of streams of traffic, and also their classification depending on type of network function.

The integrated algorithm has the following points.

Step 1. To execute identification of an arrangement of virtual network functions concerning topology of network infrastructure of virtual data center.

Step 2. To estimate quantity of the started copies of each virtual network function and to rank them in order of requirements of network infrastructure. The requiring is estimated concerning traffic flows which transferred through copies of virtual network function.

Step 3. To define a load of physical and virtual computing nodes for each copy of function.

Step 4. On the basis of the data obtained on steps 1 and 2 to make comparison of data and define virtual network functions which need to scaling, minimizing or turn off. The basic criteria used as input data to define set of virtual network functions are network records and events, data of the time of packages going through a network object, time of packet input and output, memory loading, the use of CPU, the intensity of dataflow, TTL.

Step 5. For virtual network functions which requiring minimizing or turn off to provide reconfiguration of topology on the controller of software-defined network and executed a stop and release of the occupied resources of virtual data center.

Step 6. For virtual network functions requiring scaling and creating the maximum load on infrastructure to executed an assessment of a way of dislocation. To distribute the functions which are most loaded network using a hybrid way of placement (the containers developed in the virtual machine). Less loaded, but requiring scaling network functions to transfer to an operating mode in the virtual machine.

Step 7. Provide step migration of the virtual machine and containers with containing network functions on the least loaded hardware computing nodes in data center. As part of our research, we propose a solution based on a hybrid approach and allowing seamless migration of containers within a single network space. The developed solution is based on a combination of two approaches to resource virtualization. All virtual network functions are deployed as form of a universal virtual machine. This virtual machine containing a set of containers. Each containers consist only for one specified network functions. When the network structure changing, the data on the content of the containers is replicated to the required compute node of the software-defined infrastructure of the virtual data center. This approach allows us do not interrupt the network connections in the process of changing the state of network devices.

The approach applied in the offered control algorithm of placement of virtual network functions allows to consider a way of placement and to organize work of virtual data taking into account the circulating flows of a traffic regulating at the same time quantity of the started copies of each function.

6 Conclusions

The investigation proposed a classification and identification model of virtual network functions based on the statistical properties of the flow and defined a systematic approach to the selection of the optimal set of attributes of the traffic flow. The results show that the classification of the traffic flows in cloud enable to improve the quality of service by 20-25% by reducing the response time by using virtual network functions.

In the future work we plan to explore more numbers of types of virtual network functions which used for traditional tasks, and for build non-standard solutions. Also, in the continuation of the research we are planned to investigate the work of the developed algorithmic solutions in a distributed network environment, including several geographically remote data centers.

Acknowledgments. The research work was funded by Russian Foundation for Basic Research, according to the research projects № 16-37-60086 mol_a_dk, № 16-07-01004 and № 17-47-560046, and the President of the Russian Federation within the grant for state support of young Russian scientists (MK-1624.2017.9).

References

1. *Bolodurina I. P., Parfenov D. I.* A model of cloud application assignments in software-defined storages // Journal of Physics: Conference Series — Vol. 803.
2. *Bolodurina I. P., Parfenov D. I.* Development and Research of Models of Organization Distributed Cloud Computing Based on the Software-defined Infrastructure // Procedia Computer Science — Vol. 103 — P. 569–576.
3. *Yong Li, Min Chen* Software-Defined Network Function Virtualization: A Survey. — IEEE, 2015.
4. *Ruozhou Yu, Guoliang Xue, Vishnu Teja Kilari, Xiang Zhang* Network function virtualization in the multi-tenant cloud. — IEEE, 2015.
5. *Aaron Gember-Jacobson, Raajay Viswanathan, Chaithan Prakash, Robert Grandl, Junaid Khalid, Sourav Das, Aditya Akella* OpenNF: Enabling Innovation in Network Function Control. — IEEE, 2014.
6. *Ying-Dar Lin, Po-Ching Lin, Chih-Hung Yeh, Yao-Chun Wang, Yuan-Cheng Lai* An extended SDN architecture for network function virtualization with a case study on intrusion prevention. — IEEE, 2015.
7. *Hassan Jameel Asghar, Luca Melis, Cyril Soldani, Emiliano De Cristofaro, Mohamed Ali Kaafar, Laurent Mathy* SplitBox: Toward Efficient Private Network Function Virtualization. — IEEE, 2016.
8. *Bolodurina I. P., Parfenov D. I.* Methods and algorithms optimization of adaptive traffic control in the virtual data center // 2017 International Siberian Conference on Control and Communications - Proceedings. — IEEE, 2017.