ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

An Analysis of QoS in SDN-based Network by Queuing Model

Dr. Amer Abu Jassar*

Irbed National University, Jordan[1]

E-mail: amer_abu_jassar@hotmail.com

Abstract: The analytical approach for analysis the quality of service characteristics of Software-Defined Networking is proposed in the paper. Software-Defined Networking is perspective concept that aimed to improving the performance and efficiency of computing networks. Current complexity of networks based this concept requires define the adequate methods for the calculation work characteristics. The model characterized by a Poisson arrival process with parameter λ and service time characterized by hyper-exponential distribution with r phases proposed to use for analysis QoS. Performance analysis using the proposed queuing model indicates that the model is accurate. Such QoS characteristics as maximum flow intensity that can be serving on switch and on controller, average packet sojourn time on single serve elements, delay and CP utilization of serve elements are calculated.

Keywords: Software-defined networking; Queuing theory; Controller performance; Blocking probability; Quality of service

I. INTRODUCTION

The analysis of current state and development trends in computer networks shows that the potential growth of technology performance and effectiveness for networks based on traditional technologies is almost exhausted. Many problems in functionality occur due to increasing difficulties of routing, network configuration and traffic management mechanisms [1]. The new requirements of high-speed Ethernet and data center networks especially taking into account for the future network development. The appearance of a fundamentally new approach known as Software-Defined Networking (SDN) allows expanding the possibilities of data transfer and processing networks [2-13].

The basic of SDN concept is separation data plan (physical layer) from control plane (logical layer) in computer network. Separation carried out due to transferring control functions from communication equipment routers to separate application server - controller. The control plane is responsible for computing, planning and control data transmission. The data plane is responsible for data forwarding between communication equipment according to the control messages received from controller. The interaction between the control plane and data plane protocol carries by OpenFlow protocol [10]. Implementation in existing networks SDN concepts allows increasing of quality of services, effectiveness of routing and traffic engineering; apply effective mechanisms of services virtualization and network resources distribution. However, additional network elements and modified communication equipment, for example, OpenFlow switches and controllers bring certain degradation in quality of services provision [10,11]. The increasing or changing such parameters as delay, jitter, and packet loss can adversely affect quality of service, especially on such type of services as voice and video transmitting. Evaluation of impact there parameters on quality of services is advisable to carry out by means of analytical models. The basic requirement for analytical model describing the interaction between elements in SDN-bases network is the take into account of the actual functional principles: asymmetric switches to OpenFlow controller interactions, various requests intensity and feedback. A choice of adequate mathematical model is one of the main task that should be solved for effective analyse and estimation of QoS parameters in SDN-based networks. Querying theory is one of the well-proven methods for analysing data transmission networks [9,13]. The analysis of different QoS parameters of SDN-based networks has been attempted in research works. In particular, the models of cloud computing network with various types of queuing system M/M/1/k, M/G/m/k, M/G/m proposed in [6,12]. The model of cloud data center based on queuing theory is given in [4,5]. Queuing model M/M/1 of software-Defined Networks with multi-tier applications is given in [7]. Queuing system to controller-switch interaction [8], but feedback doesn't take into account. Approach used in [3] and does not provide all type of Interaction of SDN-based network and allow to analyse only controller or switch QoS characteristics.

The goal of this work is to analysis network parameters that affect the quality of service in SDN-based networks by using queuing theory. The queuing system must be responsible to real SDN functional principles, should be able to handle any traffic intensively that going from the switch to the controller and take into account more than one switch.

The analysis of basic principles of functioning and interaction SDN discussed in Section 1. The main features of the distribution of flows from end-users to the switches and switches on the controller are described. Queuing systems that allows most fully describe the behaviour of SDN-based network proposed in Section 2. Formula for evaluation QoS characteristics as the value of delay in the transmission of applications, speed processing of applications and block probability calculation is also proposed. The experiment that allows evaluates the quality of service characteristics and analyse a properly work of proposed queuing systems are given in Section 3.

II. AN ANALYSIS OF DATA EXCHANGE MECHANISM IN SDN-BASED NETWORK

Effective functioning and providing services to customers in SDN-based networks is carrying out through the coordinated interaction of three separate architectural layers: the application plane, control plane and data plane. SDN-based network has a special type of network equipment – controller and switches supported OpenFlow protocol. The controller can be interpreting as a server, which manages the operation of Network Operating System. Like the traditional operating system, the Network Operating System manages all network resources. Over the controller operates a variety of applications that implement network services, such as routing, load balancing, a variety of protocols, gateways, firewalls, encryption, DPI, NAT, etc. In accordance with [1,2] controller receives requirements for the transmission medium, from application plane through Northbound API. OpenFlow switches transmit to the controller actual information about current characteristics of transmission medium through Southbound API. Formed control commands send to OpenFlow switches. On the basis of this commands OpenFlow switches transmit data to customers. Such type of network function organization allows optimizing the data transmission, dynamically managing the data flows rather than individual packets, load balancing, rapid control of the implementation of the information security requirements of the network. Figure 1 shows the interaction process between elements of SDN architecture.

The correct controller-switch interactions in SDN-based networks carry out by control flow. The initializing the stable channel for services providing and data transferring between two or more end nodes achieved by performing the following steps [13]:

1. Analyzing the customer's request that come to physical and virtual switch ports from different nodes with different intensively. The OpenFlow switch performs the initial processing of received requests in accordance to certain algorithm. Searching for matches between coming requests and existing requests in forwarding table is carried out. As a result, there are several possible scenarios:

1.1 If the field «Match» matches in the received request and existing in FlowTable request have been find. The request automatically forward to another devises on Data plan, for example, to OpenFlow switch.

1.2 If the field «Match» matches in received customer's request complies with existing in FlowTable, the switches generates new OpenFlow-message to the controller with information about receiving request and forward it to Control plan.

The intensity of this messages depending from using algorithms and timestamps.

2. The controller processes the arriving message and generates a response control message to OpenFlow switch.

3. The switch handles the request received from the controller and forward data according control messages.

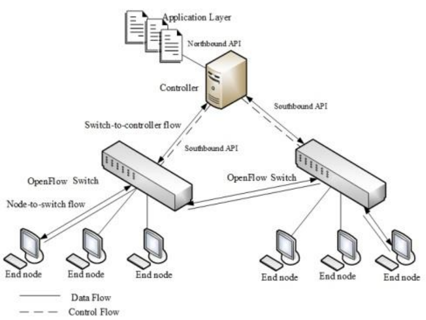


Figure 1: The scheme of control and data flow in SDN-based network.

In general, it can be noted that the interaction of physical plane complies fully reacted in a conventional SDN-based network. However, the inter-layer interaction imposes a number of additional features – existing feedback between switch and controller, different amount of new requests that arrived to switch ports, many parallels operations. Evaluation QoS for network infrastructure that include parallel and successive processes can be performed by the following formulas:

$$T(t_{\overline{a,b}}) = Min(T(t_a),...,T(t_b)) \qquad (1)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

$$W_H(t_{\overline{a,b}}) = \sum_{a \le i \le b} W_H(t_i) \quad (2)$$

$$D((t_{\overline{a,b}}) = Max(D(t_a),...,D(t_b)) \quad (3)$$

$$R(t_{\overline{a,b}}) = \sum_{j=0}^{1} f \sum_{i=a}^{b} (k(t_a) * k(2t_a - 1)R(t_a)) * ... * (k(1-t_b) + \\ + k(2t_b - 1)R(t_b)) \quad (4)$$

Where $T(t_{\overline{a,b}})$ is minimum response time for complex task $t_{\overline{a,b}}$ execution; $C_H(t_{\overline{a,b}})$ is minimum throughput for complex task $t_{\overline{a,b}}$ execution; $R(t_{\overline{a,b}})$ is network reliability, $\{t_a, t_{a+1}, ..., t_i, ...t_b\}$- subtasks that must be performed to provide service. Formula (1.4) used to evaluate reliability of SDN-based network, where $R(t_{\overline{a,b}})$ include the sum of all subtasks probabilities that execute in service providing process in this way $k(t_a)$ and $k(t_b)$ show subtasks reliability t_a and t_b respectively. The equations (1-4) allow evaluating QoS characteristics that depend on both physical and software SDN-based network components. QoS performance evaluation expediently carried out before the implementation stage by mathematical modeling. The modeling and analytical analysis help to investigate the maximum amount of data that simultaneously can be transmitted, packet sojourn time, and packet delay variation and packet loss.

III. AN QUEUING MODEL SELECTION

The main purpose of the mathematical model is the ability to fully describe the working principle of SDN-based network. It should be flexible to deal with any query volume of traffic going to the controller. In addition, the model should be easily expandable and take into account the required number of nodes. Thus, the choice of mathematical tools for modeling and subsequent analysis of the SDN-based network must include the following characteristics - the type of traffic and its intensity, the number of servers, as well as the rules for processing incoming flow of requests. The following features should consider:

- The intensity of customers' requests in SDN-based network can vary considerably. Requests can come one by one or in groups of random size;
- the characteristics and structure of data flow and processing rules for different types of services (video, speech or computing gaming services) are different;
- the customs' requests accumulated in the buffer or drop in case when the number of arriving requests longer than can be served;
- The controller-switch interaction assumes the presence of feedback between controller and OpenFlow switch.

For SDN-based network analysis should be taking into account fact that actual flows are heterogeneous: requests come from a variety of sources with different intensity. The value of the resulting flow rate at each moment depends on the load of group belonged to source (video, voice, data) and the ratio of the number of sources. Serving multiple devices can operating simultaneously with different service intensity for complex service providing Meaning the intensity of served time depends on several characteristics of OpenFlow switches and controllers. The functionality of SDN-based network can be specifying as $M / H_r / m$. This model characterized by a Poisson arrival process with parameter λ, amount of OpenFlow switches equals m, service time has Hyper-exponential distribution with r phases (H_r) with mean $1/\mu_r$.

Proposed queueing model has some limitation. Model give ability to analyze two-dimensional state-space representing the number of customers and the number of phases remained to be served for the customer in service. For simplification the phase distribution r in proposed model equals two. The type of using model is $M / H_2 / m$. Unlike [7, 9] the proposed queueing model allows realize realistic approach; take into account more than one flows between data plane (communication equipment) and feedbacks. The scheme of proposed queueing model represents on Figure 2.

The density of input request flow (M) from customers to switch can be defined by function $b(t) = \lambda e^{-\lambda t}$, the density of output requests flow from switch to controller (H_2) can be defined by function $a(t) = p\mu_c e^{-\mu_c t} + (1-p)\mu_{sw} e^{-\mu_{sw} t}$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

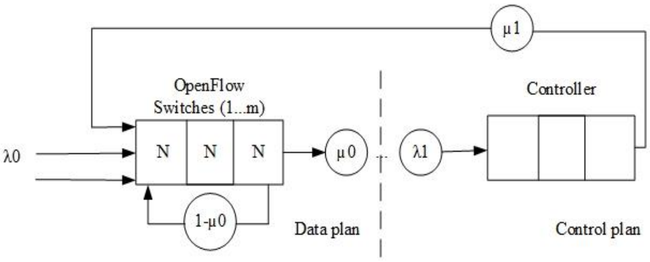


Figure 2: Proposed queuing model.

Requests from customers arrive at the switch ports (1 ... m) with different intensity λ_i. λ_0 is common requests arrival intensity from different sources $\lambda = \sum_i^k \lambda_i$. We consider a several possible scenarios:

- The input customers' requests processed on OpenFlow switch and forward from switch to controller with service intensity and probability. This case belongs to inter-layer interaction.
- The input customers' requests processed on OpenFlow switch and forward to another OpenFlow switch with service intensity and probability. In this case, requests do not leave Data plan. Is intensity of the flow of incoming requests for Control plan (switch-to-controller or controller-to-controller), is controller request service intensively.

For models type $H_2/M/m$ Laplace transform for H_2 has the next form: $A^*(s) = p\frac{\mu_0}{s+\mu_0} + (1-p)\frac{\mu_1}{s+\mu_1}$, for M has the next form: $B^*(s) = \frac{\lambda_0}{s+\lambda_0}$ [8].

The important quality of service parameters of SDN-based network like a network with high priority and real-time applications are [1,3]

- Delay and jitter;
- Controller and switches processing speed;
- Blocking probability and packets loss.

An analysis and evaluation of these QoS parameters by proposed queueing model are given in following sections.

3.1 An Evaluation of Processing Speed and Delay

The customers' requests arrive to the server's equipment, in particular, OpenFlow switch. Some requests serve immediately another redirect to the waiting place or buffer and awaiting service queue. Service waiting time depends on service discipline. The average service waiting time (W) is a major component of the delay. W is defined as the time spent in the network from the moment when packet arrives at service equipment to time when it leaves serve equipment. W in queuing system $M/H_2/m$ can be obtained by transforming the Little's low [9]. A function that determines the probability of service waiting time in the queuing system can be defined as:

$$P_{w(t)} = \sum_{i=0}^{m} p_{w_i}(t) e^{-(m\mu_i - \rho)t}, \quad (5)$$

Wher m is amount of incoming requests, ρ is the rate arrival intensity (λ_i) of request to serve intensity (μ_i), p_{wi} is i -th request awaiting probability.

Average duration of waiting in the system $W(t)$ can be calculated as follows:

$$W(t) = \frac{P_{w(t)}}{n\mu - \rho} = \frac{1}{\eta - \mu}, \quad (6)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(A High Impact Factor, Monthly, Peer Reviewed Journal)
Vol. 5, Issue 12, December 2017

Where, η is variation coefficient. $\eta=\sqrt{\frac{c_2^2}{4}+c_1-\frac{c_2}{2}}$, $c_1=\mu(\lambda_i(1-p)+\lambda_{i+1}p)-\lambda_i\lambda_{i+1}$, $c_1=\lambda_i+\lambda_{i+1}+\mu$

The absolute value of waiting time of SDN-based network where node 1 interacts with the node 2 through controller as shown on Figure 1 can be calculate as:

$$W(t)_{ab}=\begin{cases} W(t)_{sw-sw}; \\ W(t)_{sw}+W(t)_C+W(t)_{sw} \end{cases} \quad (7)$$

In this case waiting time can be written as:

$$W(t)_{ab}=\begin{cases} \frac{\rho^2}{\eta-\rho}; \\ \frac{(1-\eta)^2\mu_1^2\rho^2}{\eta\mu_0^2(\eta\mu_0-(1-\eta)\mu_1\rho)} \end{cases} \quad (8)$$

The average value of the time interval between successive requests can be calculated by the formula:

$$\bar{\tau}=\frac{1}{N}\sum_{i=0}^{N}(t_{i+1}-t_i) \quad (9)$$

Where t_i is timestamp of i-th request arriving; N is number of time intervals between the analyzed request $\overline{1,i}$.

3.2 An Evaluation of Packets Loss Probability

Assume that requests can be arrival to the entrance of queuing system in time duration equal. State system “denial of service” may occur at any point in time for any request. The probability that serve equipment not available can be calculated by arrival requests probability $P(H_i)$ [8]:

$$P_i=\frac{P(H_i)}{\sum_{i=0}^{k}P(H_k)} \quad (10)$$

Where $P(H_i)$ is current blocking probability and $\sum_{i=0}^{k}P(H_k)$ is common blocking probability, $P(H_i)$ and $\sum_{i=0}^{k}P(H_k)$ represented by exponential distribution law. In this case, formula (10) will be representing as follows:

$$P_i=\frac{\frac{1}{\sqrt{2\pi\sigma}}e^{-(i-\rho)^2/2\sigma^2}}{\sum_{k=0}^{j}\frac{1}{\sqrt{2\pi\sigma}}e^{-(i-\rho)^2/2\sigma^2}}=\frac{1}{\sum_{k=0}^{j}\exp\left|\frac{-(i-2\rho+k)(i-k)}{2\sigma^2}\right|}. \quad (11)$$

Where k is amount of receiving requests, i is current request, σ^2 load intensity variance, π equilibrium of state probability.The probability of a denial of service depends on the arrival requests intensity $\lambda_1,\lambda_2...\lambda_c$ and with condition

$\sum\lambda>mr\mu$ equal to probability when all sere devices are busy. It is called blocking or lost probability and denoted by P_d . P_d can be calculate according to the formula:

$$P_d=\frac{1}{\sum_{i=0}^{m}\exp\left[\frac{-i(-2\Lambda+m)(i-m)}{2\sigma^2}\right]}\frac{\sigma^2}{\rho} \quad (12)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

Where Λ is the network load that received during the average service time $W(t)$.

IV. EXPERIMENTAL RESULTS

Matlab + Simulink package was used for obtaining experimental of result. The fragment of designed in Simulink experimental network depicts on Figure 3. Proposed fragment assumes some limitations:

- The overall traffic arrival process at the switch and the controller is Poison. Further exponentially distributed service times are used for the switch and the controller.
- Each unknown packet/request flow sent to the controller. It is means that controller has a maximum load.
- The proposed scheme Switch to Controller has the same queuing incoming requests both at switch (Port queue), and at the controller (Port queue1). Infinite buffer is assumed at the switch, as typically it is quite large. Ordinary queue both on switch and on controller is FIFO (First input First Output).
- Serve request time on OpenFlow switch (Switch block) is restricted by "Timeout" field Serve request time on Controller is unlimited.

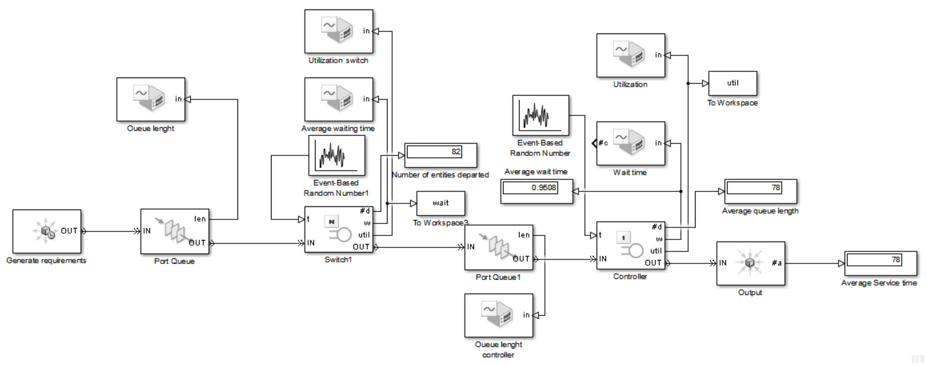
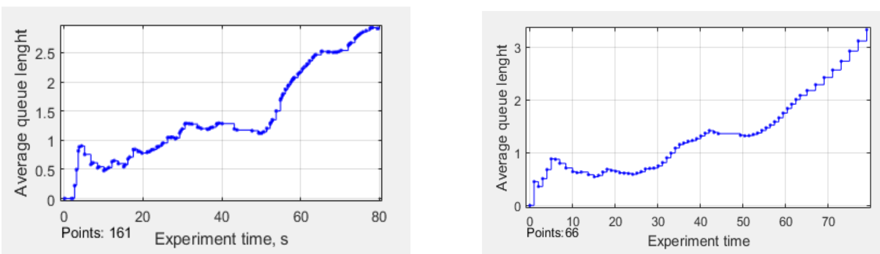


Figure 3: Experimental fragment for queuing model $M / H_2 / m$

Queue length block was used to visualize the queues on the specified ports OpenFlow switch; Average waiting time block was used as a visualizing query processing time on switch; Wait time block was used as a visualizing query processing time on controller; Utilization block was used to show the controller CPU utilization. Number of entities departure, To workspace, To workspace 3 were used to gather statistic of served requests. Block Time-Based Entity Generator was used to simulate the sequence of arriving at the input of the system service requests. Block Event-Based Random Number was used for the formation of slots used as the time of the service applications in the device.

The number of repeated experiments amounted to 20 times; the collecting statistics time is 80 seconds. The experimental results are presented in the following graphs. The average queue length of requests on switch depicted on Figure 4a and the average queue length on controller depicted on Figure 4b.



(a) **(b)**

Figure 4: Queue length on controller (a) and on switch (b).

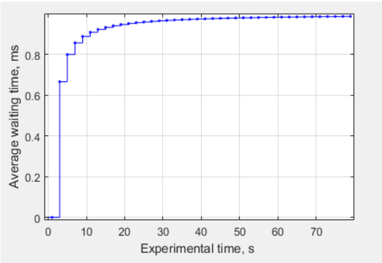
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

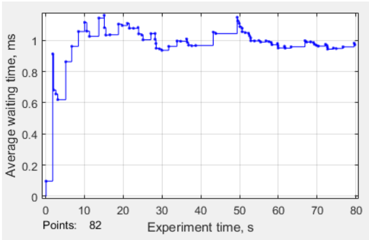
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

The obtained results show that average queue length on controller less than on single switch. The summery amount of requests (points) that arrive to the controller from all switches is 161 items. The summery amount of requests (points) that arrive to the switch from the customers is 66. The experimental results for average waiting time are depicted on Figure 5.



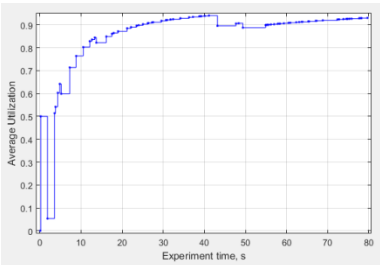
(a)



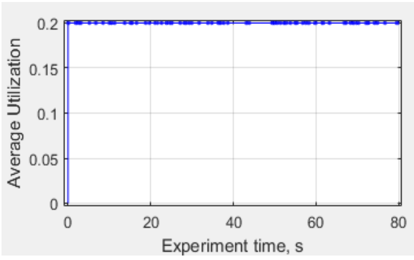
(b)

Figure 5: Average waiting time on controller (a) and on switch (b).

The controller processor utilization and single switch processor utilization are depicting in Figure 6.



(a)



(b)

Figure 6: Controller utilization (a), single switch utilization (b).

Maximum waiting time on single switch is 1,2 msec, maximum waiting time on controller less than on single switch and equal 1,0 msec. Should be noted that the average waiting time on switch has a big variation. The CPU plots shows that controller utilization for this amount of request is 20%, but utilization of single switch consist 80-90%. The meaning of average waiting time (6) and blocking probability (10) are given in Table 1.

Timestamp Interval	τ	Absolute value of waiting time $W(t)_{ab}$, ms	Absolute value of blocking probability, %
10	5.2×10^{-3}	3,514	0,0009
20	5.2×10^{-3}	3,544	0,00092
30	5.3×10^{-3}	3,04	0,0009
40	5.2×10^{-3}	2,079	0,00085
50	5.3×10^{-3}	2,033	0,00083
60	5×10^{-3}	2,384	0,0008
70	5.2×10^{-3}	2,428	0,00088
80	5.3×10^{-3}	2,844	0,00086

Table 1: Obtained QoS characteristics of proposed queuing system.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 12, December 2017

Table 1 depicted results for the segment of SDN-based network for 20 tests. The maximum number of request is 166 as obtained in modeling experiment. According to the obtained results, the feedback link and inverse requests increase the average waiting time in network fragment. The average value of waiting time bigger than common waiting time on single serve elements (Figure 5). The value of block probability belongs to the normal range. Determination boundary values of the proposed model are aim of further research.

V. CONCLUSION

An analytical model of SDN-based network based on queuing theory proposed in the work. The proposed queuing system assume that the flows are heterogeneous: requests come from a variety of sources, from customers to OpenFlow switches and from OpenFlow switches to controller, with different intensity. The formulas and shows the feedback link between switch and controller. The applicability of the proposed model is determined by next quality of services parameters average waiting time for single serve equipment both OpenFlow switch and controller, average length queue on controller and OpenFlow switch and processor utilization. The blocking probability and absolute value of waiting time including feedback was obtained by using proposed formulas. The proposed queueing model is flexible and allow including more than two OpenFlow Switches. The obtained results show that feedback reverse queries increase network load by about 7-10%, blocking probability for experimental results equals 0,00083-0,00092.

VI. REFERENCES

1. AN Bruno Astuto, M Marc, et al. A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. IEEE Communication surveys and tutorials 2014; 16: 1617-1634.
2. B Andrea, B Robert, et al, OpenFlow Switching: Data Plane Performance. IEEE International Conference on Communications 2010; 1-5.
3. CC Mun, Jean-François H, et al. On Realizing a Broadband Kernel for Multimedia Networks. Multimedia Telecommunications and Applications 1996; 56-74.
4. K Hamzeh, M Jelena, et al. Modelling of Cloud Computing Centers Using M/G/m Queues. International Conference on Distributed Computing Workshops 2011; 87-92.
5. K Hisashi, Modeling and Analysis - An Introduction to System Performance Evaluation Methodology. Addison-Wesley 1979; 18: 1-465.
6. J Michael, O Simon, et al. Modeling and Performance Evaluation of an OpenFlow Architecture. International Teletraffic Congress 2011; 1-7.
7. B Jing, Z Zhiliang, et al. Dynamic Provisioning Modeling for Virtualized Multi-tier Applications in Cloud Data Center. International Conference on Cloud Computing 2010; 370-377.
8. K Asif, D Nirav, Enabling Hardware Exploration in Software-Defined Networking: A Flexible, Portable OpenFlow Switch. International Symposium on Field-Programmable Custom Computing Machines (FCCM) 2013; 145-148.
9. K Leonard, Queueing Systems - Volume 2: Computer Applications. John Wiley and Sons, 1976; 1-576.
10. OpenFlow Switch Consortium and Others. OpenFlow Switch Specification Version 1.4.0. Open Networking Foundation 2013.
11. OpenFlow Switch Specification Version 1.1.0. 2011.
12. NC Rodrigo, R Rajiv, et al. Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments. International Conference on Parallel Processing (ICPP) 2011; 295-304.
13. O Tkachova, S Issam, et al. Mathematical models for analysis Software-defined networks. Information Technologies and Knowledge 2015; 9: 111-123.