Investigation of Queue Utilization on Network Routers by the Use of Dynamic Models

Oleksandra Yeremenko, Tetiana Lebedenko,
Tetiana Vavenko, Maxim Semenyaka
Telecommunication Systems Department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
alexere@ukr.net, tatyana_n_lebedenko@mail.ru,
tv_vavenko@mail.ru, maxisemen@gmail.com

***Abstract*—Results of investigation of impact factors to queue utilization dynamics on telecommunication network routers were presented. As was shown by the analysis, use of dynamic model, obtained by applying a Pointwise Stationary Fluid Flow Approximation (PSFFA), demonstrated that the average queue length converges to a steady state value for some time, which can range from few to tens of seconds. It is determined that the use of the steady state estimations when calculating the average queue length is possible only after the end of the transient process. Otherwise, it is advisable to use a more accurate differential model. As the results of the analysis, the duration of the transient process is influenced by such factors as the flow rate, capacity of the router interface, service discipline, etc. Within this research it was shown that the use of non-linear differential model can improve the accuracy of the calculation of the average queue length, depending on the state of the interface and the selected service discipline.**

Keywords—Quality of Service; router interface; flow rate; throughput; queue utilization; average queue length

I. INTRODUCTION

As it was shown by the analysis [1-3], the numerical values of the main parameters of Quality of Service (average delay, jitter, packet loss probability) is mainly determined by the effectiveness of solving problems of buffer resource management in telecommunication network, i.e., queues of packets organized on the routers interfaces. Whereas queue length changing contributes to packet delay variation (jitter) and queue overload in general lead to increase of the number of dropped packets. In this context, the technological mechanisms of queue management should have effective procedures to analyze the state of the router interface that lead to management decisions comply with the requirements for ensuring the required level of Quality of Service (QoS).

In scientific researches on the challenges for the organization and control queues overloading on the network routers, usually finds its application the queuing theory [4-6], with which it is possible to estimate analytically the impact of the interface state (throughput, utility, and maximum size of buffer) on average queue length, and through it to numerical value of average delay, jitter and packet loss probability. However, the tools of queuing theory allow to obtain an adequate estimation of the required parameters for a steady-state interface operation, i.e., at the end of transient processes associated with a change in the state of interface. Given the fact that the processes of estimating the state of interface and subsequent queue management are real time processes of tens milliseconds range, obtained estimations of the interface state using the limit values of probabilities may differ from the values corresponding to the dynamics of transient process in progress generally. Therefore, it is essential in the queue management to have models describing the dynamics of the interface state changes in time in order to obtain more accurate estimates of the queue length and associated QoS parameters.

II. MATHEMATICAL MODEL CORRESPONDING TO QUEUE UTILIZATION DYNAMICS ANALYSIS

There are currently known a lot of types of mathematical models, based on different approximations of the dynamics of changes in state of telecommunication network (TCN) router interface. List of approximations used to describe the dynamics of interface state changes is presented below [7]:

- Simple Stationary Approximation (SSA);
- Stationary Peakedness Approximation (PK);
- Average Stationary Approximation (ASA);
- Closure Approximation for Nonstationary Queues;
- Pointwise Stationary Approximation (PSA);
- Pointwise Stationary Fluid Flow Approximation (PSFFA);
- Modified Offered Load Approximation (MOL);
- Fixed Point Approximation (FPA).

The most efficient in relation to adequacy and clarity, in our opinion, is a model based on the use of system of nonlinear differential equations of the network state $\dot{x}(t) = dx(t)/dt$ obtained by the Pointwise Stationary Fluid Flow Approximation, PSFFA, where under the network state was understood the average queue length on the router interface. Using this model it is possible to estimate the influence of interface state, flow characteristics and packets service disciplines on the dynamics of router utilization. According to PSFFA, the flow conservation on the router interface and method of Pointwise Stationary Approximation, PSA, are combined into single nonlinear differential equation for the purpose of approximation of the non-stationary models at each time interval.

Within the chosen model there are known the following parameters: λ is the ensemble average flow rate (packets per second, 1/s) entering the analyzed queue; μ is the interface throughput (packets per second, 1/s) allocated to this queue; $\rho = \lambda/\mu$ is the queue utilization. According to the condition of flow conservation, rate of change of the average queue length at the TCN router interface will be equal to the difference of flow in and flow out [8]:

$$\dot{x}(t) = -f_{out}(t) + f_{in}(t) , \quad (1)$$

where $f_{in}(t) = \lambda(t)$ and $f_{out}(t) = \mu\rho(t)$ are rates of incoming and outgoing flows, respectively.

Then the rate of change of the average queue length at the TCN router interface takes form:

$$\dot{x}(t) = -\mu\rho(t) + \lambda(t) . \quad (2)$$

In modeling of the interface by queuing system M/G/1 (single-channel queuing system with Poisson arrival and general distribution of service time) dynamics of average queue length of the TCN router can be described by nonlinear differential equation of the form:

$$\dot{x}(t) = -\mu\left[\frac{x+1-\sqrt{x^2+2C_s^2+1}}{1-C_s^2}\right] + \lambda(t) , \quad (3)$$

where C_s^2 is the squared coefficient of variation of the service time distribution.

Depending on the value of the coefficient C_s^2 may be determined the next special cases of model PSFFA M/G/1: M/M/1 (for $C_s^2 = 1$), M/D/1 (for $C_s^2 = 0$), and M/E$_k$/1 (for $C_s^2 = 1/k$, $k \geq 1$). Thus while using in approximation different types of queuing systems (3) will take form:

- M/M/1:

$$\dot{x}(t) = -\mu\left(\frac{x}{x+1}\right) + \lambda , \quad (4)$$

- M/D/1:

$$\dot{x}(t) = -\mu\left[(x+1) - \sqrt{x^2+1}\right] + \lambda , \quad (5)$$

- M/E$_k$/1:

$$\dot{x}(t) = -\mu\left[\frac{k(x+1)}{k-1} - \frac{\sqrt{k^2x^2+2kx+k^2}}{k-1}\right] + \lambda , \quad (6)$$

where parameter k denotes the number of service stages.

III. Investigation of Queue Utilization Dynamics on the Router Interface

With the selected model (1)-(6) was performed the analysis of transient processes corresponded to queues utilization on the telecommunication network router in terms of the influence of the interface state on the duration of these processes and accuracy of estimates of the average queue length. It is obtained that time of convergence is influenced by following factors: interface throughput, flow characteristics (intensity, packet size), and the type of packets service discipline. Taking into account these factors was carried out with the use of the following queuing systems: M/M/1, M/D/1 and M/E$_k$/1, which accurately describe real flows in modern multiservice TCN [8].

For illustration represented the next example. Suppose that the VoIP call stream need to be transmitted. Let us consider that it is used the ITU-T G.711 codec for audio companding. Some codecs use compression in order to reduce the bandwidth required for a VoIP call, which inevitably results in the loss of detail of the original signal, hence in general the better the call quality required, the more bandwidth that will be required per call. VoIP codecs generally produce a constant bitrate stream, with bandwidths as shown in Table I [9].

TABLE I. ITU-T G.711 VoIP Codec Characteristics

ITU-T Codec	Codec Type	Bitrate (bps)	IP Packet Size (bytes)	IP Bandwidth (bps)
G.711	PCM[a]	64000	120	96000
G.711	PCM	64000	200	80000
G.711	PCM	64000	280	74659

a. Pulse Code Modulation, PCM.

Results of investigation performed are shown in Table II, where queuing systems M/M/1, M/D/1 and M/E$_k$/1 analyzed for different packet sizes and consequent IP bandwidth. Queuing utilization dynamics was studied according to different number of flows: one VoIP stream, two and three concurrent VoIP streams. Parameter characterized the dynamics is the time of convergence of the average queue length to its limit value.

TABLE II. Investigation of Queue Utilization Dynamics for VoIP Flows

Queuing System		M/M/1	M/D/1	M/E$_k$/1
Packet Size 120 B				
Queue Utilization	*Number of Flows*	*Convergence Time, s*		
ρ=64/96	1	0.4	0.28	0.42
ρ=128/192	2	0.2	0.13	0.17
ρ=192/288	3	0.16	0.09	0.13
Packet Size 200 B				
Queue Utilization	*Number of Flows*	*Convergence Time, s*		
ρ=64/80	1	2.1	1.3	1.76
ρ=128/160	2	1.3	0.65	0.87
ρ=192/240	3	0.9	0.47	0.58
Packet Size 280 B				
Queue Utilization	*Number of Flows*	*Convergence Time, s*		
ρ=64/75	1	6.3	3.4	4.8
ρ=128/149	2	3.5	1.9	2.67
ρ=192/224	3	2.8	1.2	1.83

Within the analysis by the average flow rate λ understood the bitrate of the VoIP stream, and μ is an IP bandwidth consumed by this flow both transformed into units of 1/s. Then it was calculated the queue utilization as $\rho = \lambda/\mu$, which varies for different sizes of packets in

VoIP flow and equal to 0.67, 0.8, and 0.86 for 120 B, 200 B, and 280 B packet size, respectively.

On Fig. 1-3 convergence analysis of average queue length shown for the case of packet size equal to 120 B for different queuing models of router interface. This case considered to be the best in terms of convergence time according to obtained results (see Table II) for all packet sizes possible in G.711 standard VoIP call streams.

As it can be seen from Fig. 1, with initial data λ =66.7 1/s and μ =100 1/s if the interface is modeled by queuing system M/M/1, the value of the average queue length converges to its limit value (2 packets) with different number of VoIP streams transmitted at different rates. Namely, when ρ=64/96 and one VoIP stream transmitted convergence time equals to 0.4 s, while at ρ=128/192 it is 0.2 s, and at ρ=192/288 it is 0.16 s with two and three concurrent streams transmitted, respectively. Which means that the greater number of concurrently transmitted flows, the less time needed to convergence of average queue length to its limit value, i.e. convergence time is reduced by 2.5 times even during transmission of three VoIP concurrent streams. Thus, flow aggregation has the positive impact on the dynamics of changes in router interface queue utilization. However, if the queue management is performed with a period of 100 ms, for example, the average queue length is approximately 0.2 packets that is significantly (up to 10 times) differ from the average queue length of 2 packets in the steady state.

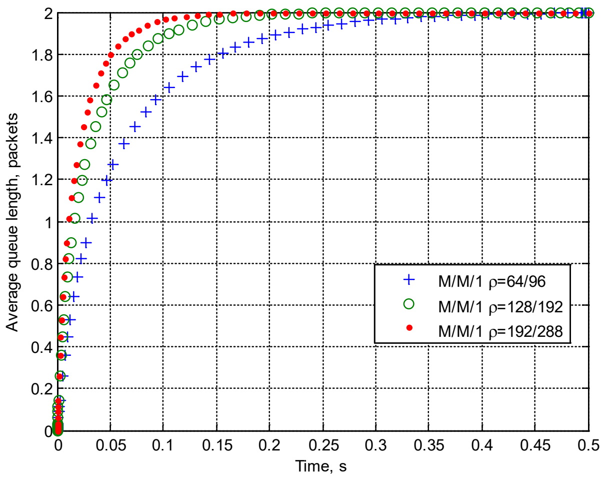

Fig. 1. Convergence analysis of average queue length to the limit value for M/M/1 queuing system with packet size 120 B.

With the same initial data λ and μ if the interface modeled by queuing system M/D/1 (Fig. 2), the average queue length converges to its limit value (1.33 packets) for different number of transmitted VoIP streams at different rates too. That is to say, when ρ=64/96 and just one stream transmitted convergence time is equal to 0.28 s, while at ρ=128/192 it is 0.13 s and at ρ=192/288 it is 0.09 s with two and three concurrently transmitted streams, respectively. Whereby with transmission of three streams the convergence time is reduced by about 3 times, and the average queue length at an interval of 100 ms control will be at about 0.1 packet, which is more than 13 times less than the average queue length in the steady state.

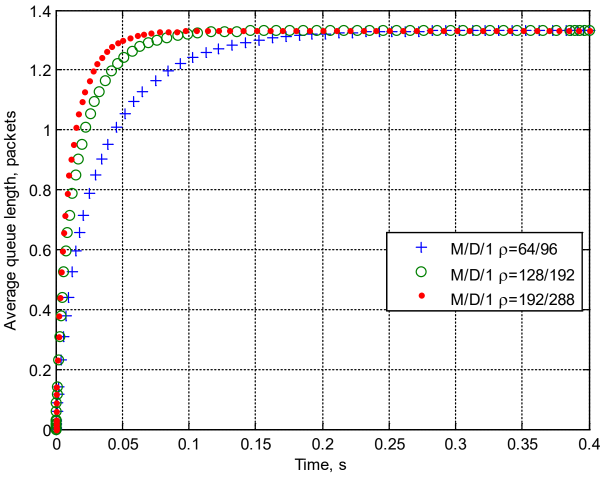

Fig. 2. Convergence analysis of average queue length to the limit value for M/D/1 queuing system with packet size 120 B.

In modeling the router interface through the queuing system $M/E_k/1$ (Fig. 3) also used the same initial parameters λ and μ as before. The value of the average queue length converges to its limit value of 1.68 packets. When one stream transmitted and ρ=64/96, the convergence time obtained is 0.42 seconds, and when ρ=128/192 with two transmitted streams, then we have convergence time as 0.17 s, and with transmission of three concurrent streams and ρ=192/288 this period is equal to 0.13 s. It should be noted that with the transmission of three VoIP call streams convergence time is reduced by roughly 3.2 times, and the average queue length at an interval of 100 ms control will be about 0.2 packets, which differs from the steady state value of the average queue length for more than 8 times.

As it can be seen the queuing model $M/E_k/1$ considering packet size of 120 B shows results similar to M/M/1 model of router interface, but in relation to convergence times the M/D/1 model demonstrates the better results and lesser periods of transient process to the steady state. Moreover, the same tendencies observed for modeling with packet sizes 200 B and 280 B.

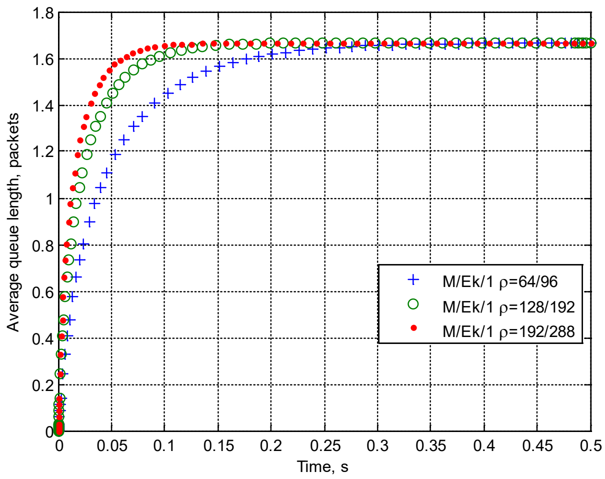

Fig. 3. Convergence analysis of average queue length to the limit value for $M/E_k/1$ queuing system with packet size 120 B.

Based on the investigation, can be distinguished the next main factors of influence to the time of convergence of the average queue length to the limit value in steady state:

1. It has been shown that a significant influence on the convergence time of average queue length to the limit value has a packet size within the transmitted flow. And there is the minimum convergence time observed with a smaller packet size. In this case, the queue utilization of the router interface had the smallest value among examined.
2. It should be noted the impact of the distribution law of flows service time on the network router interface. Among the used queuing models the best performance has a model M/D/1, compared with the models M/M/1 and $M/E_k/1$, with the same initial data (bitrate and IP bandwidth consumed). In the case of using M/D/1 with the lesser value of queue utilization convergence time of the average queue length is of the order 0.1-0.3 s, while the other models have a longer transient process to the steady state during 0.13-0.42 s. In general, with different packet sizes, and accordingly, varying queue utilization, using the model M/M/1 observed the longest transient process (up to 6.3 seconds).
3. The study showed that the aggregation of data flows (VoIP streams in represented research) allows reducing the duration of the transient process, despite the fact that queue utilization on the router interface does not change its value. It was demonstrated that the convergence time of average queue length to its limit value with increasing number of concurrently transmitted flows from one to three decreases on average in 2.5-3 times.

IV. Conclusion

Due to the high dynamics of changes of the interface state in the estimation of the average queue length and the related key indicators of Quality of Service in a number of important cases it is required the use of dynamic model, obtained by applying a pointwise stationary fluid flow approximation PSFFA and representation of state by the nonlinear differential equation, where the state was understood as the average queue length.

As was shown by the analysis, the analytical calculation of the average queue length at the router interface is an important step addressing to the main tasks of queue management. The higher accuracy of calculation of the average queue length, the more informed decisions on the volume of discarded packets from it, and the more accurately, in turn, can be predicted the value of the average delay and packet loss probability at the router interface. Using expressions to calculate the average queue length for the steady state often gives a very rough approximation of the true values. While using dynamic nonlinear model was able to show that the average queue length converges to a steady state value for some time, which can range from few to tens of seconds.

Based on the obtained results, it is determined that the use of the steady state estimations when calculating the average queue length is possible only after the end of the transient process. Otherwise, it is advisable to use a more accurate differential model. As the results of the analysis, the duration of the transient process is influenced by such factors as the flow rate, capacity of the router interface, service discipline, etc. And it was found that the higher the interface throughput, the duration of the transient period is lower. From the set of considered service disciplines it was found that in the case of using the model M/D/1 average queue length convergence to the limit value in the steady state is faster than for other models. Within this research it was shown that the use of non-linear differential model can improve the accuracy of the calculation of the average queue length, depending on the state of the interface and the selected service discipline.

Besides, the main factors of influence to the time of convergence of the average queue length to its limit value in steady state were concluded. It was shown that the smaller packet size flows demonstrate the faster transient process. Moreover, with the same initial data appropriate choice of queuing model can reduce the time of convergence. Within the research it was shown that the aggregation of data flows allows reducing the duration of the transient process with the same value of queue utilization on telecommunication network router interface.

References

[1] A.K. Kloth, *Advanced Router Architectures*. CRC Press, 2005.

[2] M. Marchese, *QoS Over Heterogeneous Networks*. John Wiley & Sons, Ltd, 2007.

[3] T.-Y. Tsai, Y.-L. Chung, and Z. Tsai, “Introduction to Packet Scheduling Algorithms for Communication Networks”. *Communications and Networking*, Jun Peng, 2010.

[4] O. Lemeshko, and M. Semenyaka, “Research of Queue Management Model That Includes Congestion Avoidance Mechanism,” in *First International IEEE Conference Problems of Infocommunications. Science and Technology*, pp. 42-44, October 2014 [Proceedings of PICS&T-2014, Ukraine, p. 203, 2014].

[5] O. Lemeshko, M. Semenyaka, and O. Simonenko, “Researching and designing of the Dynamic adaptive queue balancing method on telecommunication network routers,” in *XIIth International Conference The experience of designing and application of CAD systems in microelectronics*, pp. 204-207, February 2013 [Proceedings of XIIth CADSM`2013, Ukraine, p. 480, 2013].

[6] O. Lemeshko, A.S. Ali, and M. Semenyaka, “Results of the Dynamic Flow-Based Queue Balancing Model Research,” Modern Problems of Radio Engineering, Telecommunications and Computer Science, pp. 318-319, February 2012 [Proceedings of the international Conference TCSET’2012, Ukraine, p. 552, 2012].

[7] K.A. Alnowibet, Nonstationary Erlang Loss Queues and Networks. PhD Thesis, North Carolina State University, 2004.

[8] W.-P. Wang, D. Tipper, and S. Banerjee, “A Simple Approximation for modeling Nonstationary Queues,” in *Proceedings of the Fifteenth Annual Joint Conference of the IEEE Computer Societies (INFOCOM '96)*, vol. 1, pp. 255–262, 1996.

[9] J.W. Evans, and C. Filsfils, *Deploying IP and MPLS QoS for Multiservice Networks*, 1st Edition. Morgan Kaufmann, 2007.