Analytical and Simulation Models of Dynamic Priority Service System (DPSS)

Fatholah Dadgar Arablu[1] ID

[1]Baku State University, AZ 1148, Baku, Av. Academy of Z. Khalilova Baku, Azerbaijan

Abstract: In Differentiated Services (DiffServ) architecture, Priority Service System (PSS) plays a vital responsibility to provide Quality-of-Service (QoS) for the applications based on networks. Priority queuing systems has always been a subject of interest for analytical modelling and evaluating the performance. However, previous works have mainly focused on performing priority queuing under range dependent traffics, namely Short Range Dependent (SRD) or Long Range Dependent (LRD). Recent studies revealed that realistic traffic demonstrates a heterogeneous nature for the modern networks that provide multiple services. In this paper, the results of analytical and simulation models of dynamic priority service systems is reviewed to study the heterogeneous traffic impact on designing and performance of the systems based on networks. In this paper, the results of analytical and simulation models of dynamic priority service systems is reviewed to investigate the impact of heterogeneous traffic on the design and performance of network-based systems. Here, the time of presence of requests in limited and unlimited buffer system is restricted. If the restriction is violated, the requests will be lost. The results of experiments conducted on both models are compared and it is identified that they differ in the allowable limit (2-9) %.

Keywords: Service systems, service process, analytical and simulation model, request, transact, dynamic priority, limited and unlimited buffer.

1. Introduction

Significant advancements in networks and technologies extending communication resulted in popularising network-based computing as a paradigm to realise a computing scheme that is both cost-effective and delivers high levels of performance. In the recent years, with the focus on global-scale systems, some novel computing schemes such as Internet-based computing and grid computing were proposed. The ability to employ systems in global scales in crucial in a multiple set of domains, such as financing, industry, military, and telecommunications. Procuring Quality-of-Service (QoS) in various applications regarding networks has become an increasingly pressing demand. The architecture models for Differentiated Services (DiffServ), as an efficient scheme to provide QoS, categorises data packets to a few aggregated flows or classes to be handled according to the differentiated priorities. Hence, preparing scheduling policies considering the priorities plays a significant role in implementing DiffServ models, and impose crucial implications on their efficiency [1-4]. The study of two options – limited and unlimited queuing service systems in the service for requests with limited time for the presence in the service systems are reviewed. If the time restriction for requests' presence in the system is violated, the requests will be lost. These situations occur practically in service processes of different technical systems. Therefore, the analysis of possible situations leading to the loss in such systems is one of the important issues. The conducted studies show that although a number of research works have been dedicated to dynamic priority service

processes, it has not been possible to make the total time passed in service T_{S^*} shorter than the allowed limit as a preventive measure. The use of the algorithm that can better organize service process enables reducing the total loss of request and increases the productivity of the system. The investigations show that the solution is possible by choosing the optimum service method among the options of interaction of queue and service algorithms. It is assumed that there is a homogeneous stream of requests with a priority in the service. Dynamic priority of requests in the service process can vary depending on the situation. Service time of all requests is distributed with the same rule and it is possible for the situation to vary in two places –in queue and service during the system operation. The presence of requests in the system consists of two phases, waiting and service. The total time T_{S^*} of the presence of request in the system should not exceed T_{S^*}. The removal of the requests in the waiting phase comprises the first type of loss in the system according to a certain rule in compliance with the algorithm of the queue. According to the algorithm of the organization of service, if the time of presence of requests in the system exceeds T_{S^*} the loss of requests removed from the system on a certain rule comprises the second type of loss in the system. In this paper, four cases covered by two options of limited and unlimited queuing service (QS) models with the same type of request and facilities in N number are reviewed in the introduction to these models. In the first case, the value of the presence of requests in the system (in queue and service) and the value of the presence of requests in the system exceeding the possible maximum value (at the latest, after the removal of request from the system or service time) are determined. In the second case, the time value for the presence of requests in the system (in queue and service) and the time value for the presence of requests in the system exceeding the possible maximum value (in queue and at the latest, after the removal of request from the system (in service)), i.e. the moment of the occurrence are determined. A basic dynamic priority service system is shown in Fig. 1.
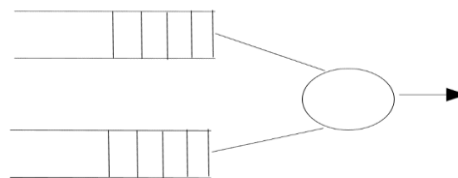
Fig. 1 Dynamic priority service system with two classes of traffic.

Limited-unlimited queuing services are shown in Fig. 2 and 3.
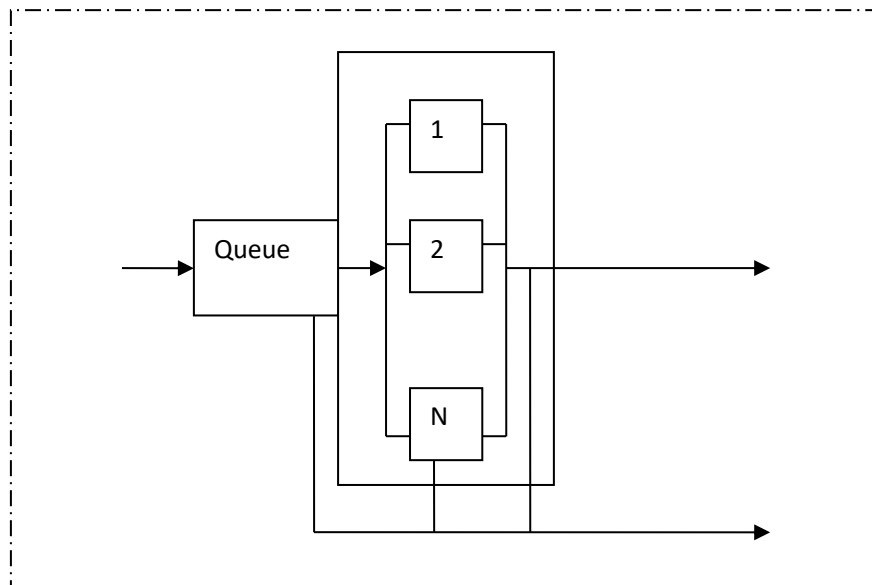
Fig. 2 limited queuing service.
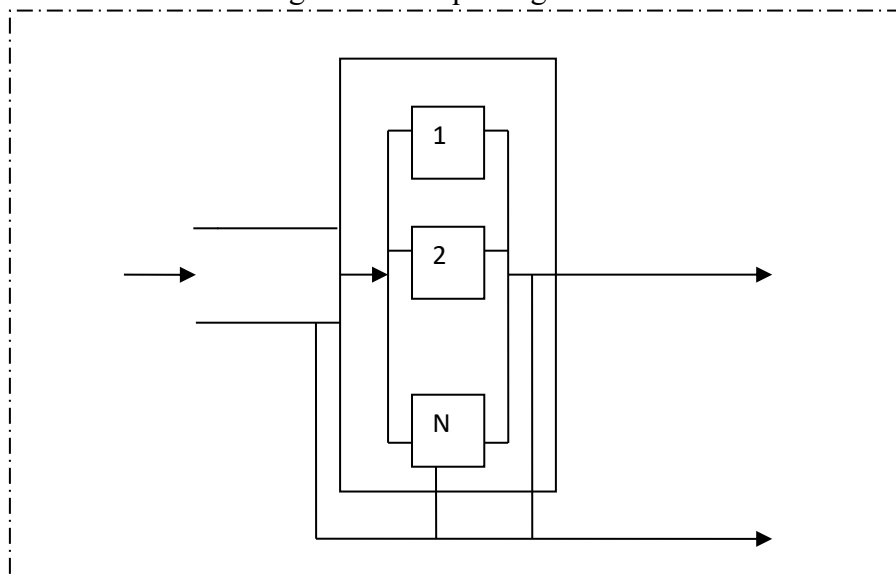
Fig. 3 Unlimited queuing service.

The remainder of this paper is organized as follows. In Sect. 2, the related previous studies carried out on the subject of dynamic priority service systems are presented. Sect. 3 discuss our proposed DPSS in detail. Moreover, the parameters employed in evaluating the performance and simulation results are introduced in Sect. 4. Finally, conclusions are provided in Sect. 5.

2. Related Works

There are numerous research in literature studying performance evaluation and analysing the mechanism of the priority service system. In the following section, the classification of the previous works are provided.

2.1. Short Range Dependent (SRD) Traffic Method

The original focus of previous researches were on queuing systems according to priority prone to Short Range Dependent (SRD) traffic [5–8]. For example, Choi et al. [5] investigated a queuing system in which traffic flows with high priority and low priority are modelled using Markov Modulated Poisson Process (MMPP), and ordinary Poisson process, respectively. From this, they successfully concluded the steady-state joint generation function of the queue length for each traffic flow in theory.

Mazzini, et al. [8] proposed a queuing system based on considering high and low priority traffic flows to be similar statistical Bernoulli distribution forms. Moreover, they studied steady-state queue length distribution and average queue length, and derived analytical expressions for them.

For priority queuing systems with traffics of the Gaussian form, Nannersalo and Norros [7] produced approximations that are applicable to queueing length distribution of the mentioned systems. The asymptotics of packet loss and delays in such systems is further studied in Mandjes et al. [6].

2.2. High quality and high time-resolution measurements

As is confirmed in multiple studies conducted recently on measurements for high quality and high time-resolution [9–13], in a variety of networks, traffic demonstrates considerable frailness over a vast range of time scales and in various networks. By employing statistically self-similar processes with substantially different theoretical properties compared to the conventional SRD processes, this network traffic fractal-like behaviour can be modelled significantly better.

2.3. Self-similar phenomenon of traffic

The self-similar traffic phenomenon is present in a wide range of networks, such as local-area networks [10], wide-area networks [12], World Wide Web [9], wireless networks [11, 13], and Variable Bit-Rate (VBR) video systems [14]. Therefore, studies concerned with priority queuing are focused on systems with Long Range Dependent (LRD) self-similar traffic [15–17]. For example, Ashour and Le-Ngoc [15] utilised Multiscale Wavelet Models (MWM) to specify a priority queuing system LRD input traffic, and examined the analytical estimations for high and low priority traffic queue length distributions.

2.4. Measurement-based method

A measurement-based method to estimate the probability of buffer overflow of queues in a priority queuing multiplexer is developed in Quan and Chung [17].

In Iacovoni and Isopi [16], a queuing system with the high priority traffic being asymptotically self-similar and the low priority traffic being precisely self-similar, is analysed. In this paper, a lower bound for the overflow probability in low priority queue is obtained.

However, these studies conducted earlier considered only priority queuing systems with homogeneous traffic. To elaborate, studies consider either SRD traffic or LRD. However, monitoring the traffic within modern networks providing multiple service simultaneously demonstrates a

heterogeneous nature. To our best of knowledge, there is hardly any analytical model that takes heterogeneous SRD and LRD traffic presence in priority queuing systems into account. To provide a solution to overcome this shortage, in this paper a novel analytical model for systems subject to heterogeneous LRD self-similar traffic and SRD Poisson traffic is proposed, which is based on Large Deviation Principles (LDPs).

2.5. Static priority approach

Previously a number of adaptive schemes based on static priority and round robin were proposed to overcome the drawbacks for each scheme.

In the model proposed by Kleinrock, instantaneous priority is dependent on a parameter with varying nature. In this model, p classes with a parameter b'' associated with them ($0 \leq b1 \leq b2 \leq \cdots \leq bp$) are considered. In this formulation, the priority of a customer with class i, arrived at time T, is given by the equation $(t - T) * bi$. In Lim et al. (1988), Lim and Kobza proposed a scheme called head-of-line priority with jumps (HOL-PJ). In this scheme, a model with p traffic classes is considered. If $i \leq j$, it is said that the class i has non-pre-emptive priority over class j. However, there is an upper limit on the time for each customer, which can be spent on a given queue. When the imposed time limit is exceeded, the customer moves to the end of the next queue with the highest priority [18].

Ozawa studied a doubled-queued system, in which exhaustive services are received by the high priority queue, while the low priority queue service is K-limited [19].

Lee and Sengupta [20] proposed and analysed a double traffic-class model in an ATM network.

2.6. Round-robin approach

In his approach, system service is carried out using the round-robin service discipline among the classes. For either classes, it is possible to define a threshold denoted by L. If a class exceeds the threshold in any of its queue lengths, the cells that are exclusively from that class are going to be serviced, until the queue length once again falls below the threshold, which is then reverted to the round robin. Performing analysis on coupled queuing systems, such as the instances described above, generally involves transform-based analysis, often leading to numerical solutions. However, obtaining a closed-form solution without simplifying the system behaviour approximations is difficult [20].

2.7. Multiple finite source queuing model

A multiple finite source queuing model possessing dynamic priority scheduling is solved in Tosirisuk and Chandra [21]. In this model, there are assumed to be P (P > 2) users for each computer system, and the requests are transmitted to the CPU to be processed. Plus, the system consists of a single CPU. A response is transmitted back to the user, once processing of a selected process is completed. The user then generates another CPU request. The considered system might be a multi-echelon system, possessing multiple bases, a centralized repair facility with a single repairperson, and no spares. What differentiates their study from other authors is the lack of pre-emptive dynamic priority service discipline. To solve this system, an approximation is derived. Simulation results demonstrate the accuracy of the proposed algorithm.

2.8. Other models

Management of distributed networks, which possess dynamic priorities, their requirements, and by considering buffer limitations is studied in [22]. This study revealed that the total loss lower limit was attained by the value specified by the queue length. The service process optimal characteristics in a distributed computer network, which complicated the performance of the operating system, was determined using these results. However, developing more advanced algorithms to organise the service process, reduces the requirements loss, and increases the network services performance. A notable feature of employing this approach is minimising the loss of claims without the service permission, due to the preventive removal of some. To study the performance of such networks, these systems can be modelled as a queuing system (QS). To solve the parameter optimisation problem, it is necessary to determine the nature of preventive removal requirements, resulting in loss minimisation.

To confirm the sufficiency of the analytical model developed in [23], simulation models should be developed for such networks and services according to certain criteria, the results should be compared for both models. What is apparent is that meeting the predefined quality of service levels in switches requires employing high-speed multimedia networks with buffer spaces consisting shared buffering, with numerous spatial priorities such as space priorities, objectified construction procedures in various applications through utilising time priority buffers, and defining the necessary rules to select the type and appearance of the s buffers [6]. Therefore, in classic priority service scheme, specifically Go type, it is generally assumed that a certain application possess higher priority compared to other bids in both species. The analysis demonstrated that on theory, the cat considered schemes with various levels of spatial priorities to control sound in lyayuschie intensities (probably) in different applications, loss of time, etc., and priorities affecting their time of s and latency in the buffer.

In [24, 25] various schemes were proposed to spatially define priorities to find the optimal size employed for Buffer Memory (in some sense) used in the nodes of different service networks.

In [26 and 27], itis proposed employing time-varying dynamic priorities from the first high-speed multimedia service network, based on its temporary functional conditions, as temporary priority.

3. The proposed approach

In this section, a DPSS method is designed by addressing the analytical and simulation models of dynamic priority service system.

3.1. Formulation and Solution of the Problem

Four cases covered by two options of limited and unlimited queuing service (QS) models with the same type of request and facilities in N number are reviewed in the introduction to these models. In the first case, the value of the presence of requests in the system (in queue and service) and the value of the presence of requests in the system exceeding the possible maximum value (at the latest, after the removal of request from the system or service time) are determined. In the second case, the time value for the presence of requests in the system (in queue and service) and the time value for the presence of requests in the system exceeding the possible maximum value (in queue and at the latest,

after the removal of request from the system (in service)), i.e. the moment of the occurrence are determined.

3.2. The service for requests and verification of the condition

The abbreviation is shown in Table 1.

Table 1 The abbreviation.

Parameter	The parameters meaning
p_1	probability of loss of resources in the waiting phase
p_2	probability to remove requests when providing services at the server
λ	Intensity of requests
μ	Severity of service time
ρ	Increasing demands
L_q	Queue length
L_s	Average number of requests in the system
τ_q	Lending queue waiting time
τ_s	Average time spent on requests in the system
m	Maximum waiting queue
τ_s^*	Average maximum possible time spent on requests in the system

Four $\tau_s \leq \tau_s^*$ are fulfilled in different computer systems, (it is consistent with the first case of both variants) and the condition $\tau_s \leq \tau_s^*$ is not met, there is no chance for the request to keep the service. Such situation can occur in the first case of both options (there is no chance for requests to determine τ_s) If service for requests and verification of the condition $\tau_s \leq \tau_s^*$ are fulfilled in the same computer, the request remaining in the system more than τ_s^* time leaves the system without waiting for the completion of the service. Such a situation can occur in the second case of both options (there is a chance to determine τ_s during the service). The above-mentioned features of service process significantly affect its analytical study. In a special case, the value of the second type of loss for the model expressing the first case of the first option does not affect the characteristics of queue and service, so that, the requests is fully serviced in any case. However, the value of the loss of second type does not affect the features of queue for the model expressing the second case of the first option. Therefore, its analytical study is more complex than the model expressing the first case of the first option.

In this case, the removal of requests, which have not been served until the end in the system (the second case of the first option), accelerates the passage of requests through the system (in comparison with the first case of the second option).

In addition, this enables using the statistic results obtained for the first case of the first option and assessing the relevant quantities in the second case of the first option. The first and second cases of the second option have an endless queue, and the removal of request from the queue is executed on the criteria expressed with the time of presence of requests in the system (there is a chance to determine the value of τ_s quantity in the service). The organization of interaction in terms of

effectiveness in the phases of queue and service of service system in the reviewed options of the model allows writing different options. In addition, it should be noted that the procedure for the removal of requests from the system should be corrected depending on the level of degradation of the system for each case of the organization of interaction of queue and service.

"The oldest" request is removed from the queue in all options of the system reviewed in models (if all places are occupied), so that, exactly this request in the queue has no chance to meet the time restriction for the presence in the system more than others. It was considered for simplicity that requests have been in the queue by the order of entry, i.e. the request, which is always placed in the beginning, is removed from queue. It should be noted that in all cases, there is the finite value of the length of queue minimizing the total loss in the wide range of the rules for the selection of requests for the service and service time with Poisson input stream and exponential distribution law. The characteristics of the reviewed QS have been determined by an analytical method with GPSS (General purpose simulation) modeling language.

4. Performance evaluation

In this part, performance of the DPSS proposed method is proposed in the problems of waiting time, length of the queue on realizations, length of the shift on the realization and limited and unlimited queue service processes.

4.1. Performance metrics

Extensive simulations are conducted to evaluate the effectiveness and performance of DPSS and compare it with analytical (A) and simulation model.

4.2. Simulation results and analysis

DPSS approach in the GPSS is implemented on Linux-Fedora 10. Simulations are performed on the proposed method and it is found that proposed method performance has better results in the problem of waiting time, length of the queue on realizations, length of the shift on the realization and limited and unlimited queue service processes.

4.2.1. Simulation results

GPSS World Simulation Report is shown in Table. 2, 3, 4, 5 and 6.

Table 2 Output of GPSS Simulator 4.2.1.

_Start_time	_End_time	_Block_	_Facilities_	_Storages_				
0.012	142.120	16	2	1				
Facility	_Entris_util_	_AVG_	_TimeAvail_	_Owner_	_Pend1_	_Enter_	_Retry_	_Delay_
_Sys	1	0.935	0.0061	1	1	1	1	1
USER2_a	CHAIN2_a	SIZE_a	RETRY2_a	AVE2_a	CONT2_	ENTR2IES	MAX2	AVE.TIME
LINE	3	1	0.00007	4	5	4	2	121.152
CECC	XNN	PRII	M11	ASSEMM CURRENT NEXT PARAMETER				VALUE
8	2	338.769	8	14				6
FEC _XN_	_PRI_	_BDT_	_Asem_	_Caren_next_par_meter _				_Value_
7	2	479.643	7	2				2

Table 3 Output of GPSS Simulator 4.2.2.

_Start_time	_End_time	_Block_	_Facilities_	_Storages_				
0.012	142.120	16	2	1				
Facility	_Entris_util_	_AVG_	_TimeAvail_	_Owner_	_Pend1_	_Enter_	_Retry_	_Delay_
_Sys	1	0.935	0.0061	1	1	1	1	1
USER2_a	CHAIN2_a	SIZE_a	RETRY2_a	AVE2_a	CONT2_	ENTR2IES	MAX2	AVE.TIME
LINE	3	1	0.00007	4	5	4	2	121.152
CECC	XNN	PRII	M11	ASSEMM CURRENT NEXT PARAMETER				VALUE
8	2	338.769	8	14				6
FEC _XN_	_PRI_	_BDT_	_Asem_	_Caren_next_par_meter _				_Value_
7	2	479.643	7	2				2

Table 4 Output of GPSS Simulator 4.2.3.

_Start_time	_End_time	_Block_	_Facilities_	_Storages_				
0.012	142.120	16	2	1				
Facility	_Entris_util_	_AVG_	_TimeAvail_	_Owner_	_Pend1_	_Enter_	_Retry_	_Delay_
_Sys	1	0.935	0.0061	1	1	1	1	1
USER2_a	CHAIN2_a	SIZE_a	RETRY2_a	AVE2_a	CONT2_	ENTR2IES	MAX2	AVE.TIME
LINE	3	1	0.00007	4	5	4	2	121.152
CECC	XNN	PRII	M11	ASSEMM CURRENT NEXT PARAMETER				VALUE
8	2	338.769	8	14				6
FEC _XN_	_PRI_	_BDT_	_Asem_	_Caren_next_par_meter _				_Value_
7	2	479.643	7	2				2

Table 5 Output of GPSS Simulator 4.2.4.

_Start_time	_End_time	_Block_	_Facilities_	_Storages_				
0.012	142.120	16	2	1				
Facility	_Entris_util_	_AVG_	_TimeAvail_	_Owner_	_Pend1_	_Enter_	_Retry_	_Delay_
_Sys	1	0.935	0.0061	1	1	1	1	1
USER2_a	CHAIN2_a	SIZE_a	RETRY2_a	AVE2_a	CONT2_	ENTR2IES	MAX2	AVE.TIME
LINE	3	1	0.00007	4	5	4	2	121.152
CECC	XNN	PRII	M11	ASSEMM CURRENT NEXT PARAMETER				VALUE
8	2	338.769	8	14				6
FEC _XN_	_PRI_	_BDT_	_Asem_	_Caren_next_par_meter _				_Value_
7	2	479.643	7	2				2

Table 6 Output of GPSS Simulator 4.2.5.

_Start_time	_End_time	_Block_	_Facilities_	_Storages_				
0.012	142.120	16	2	1				
Facility	_Entris_util_	_AVG_	_TimeAvail_	_Owner_	_Pend1_	_Enter_	_Retry_	_Delay_
_Sys	1	0.935	0.0061	1	1	1	1	1
USER2_a	CHAIN2_a	SIZE_a	RETRY2_a	AVE2_a	CONT2_	ENTR2IES	MAX2	AVE.TIME
LINE	3	1	0.00007	4	5	4	2	121.152
CECC	XNN	PRII	M11	ASSEMM CURRENT NEXT PARAMETER				VALUE
8	2	338.769	8	14				6
FEC _XN_	_PRI_	_BDT_	_Asem_	_Caren_next_par_meter _				_Value_
7	2	479.643	7	2				2

4.2.2. Analysis

A Comparative Study of the dependence of the numerical results is carried out.

$m = 10$

$\tau_s^* = 100$

$(\lambda = 0.001 - 0.005, \mu = 0.010 - 0.022)$

$\rho = 0.1 - 0.9$

$L_q = f(\rho), \tau_q = f(\lambda)$

Exponential service can be set as Equation (1):

$$L_q = \frac{P\,\rho}{(N - \rho)} \tag{1}$$

Depending on the nature of the system object, the approximations   of Equation (2) and (3)can be made:

$$\rho \ll 1 \qquad \text{when} \qquad L_q \to \frac{\rho^{N+1}}{N^2} \tag{2}$$

$$\frac{\lambda}{\mu N} \to 1 \qquad \text{when} \qquad L_q \to \frac{\rho}{(N-\rho)} \tag{3}$$

The timetable of the requirements for known requirements can be determined by the timing of the waiting time, the time spent in the system and the number of expected requirements in the system Ls as Equation (4) and (5):

$$\tau_q = \frac{L_q}{\lambda} \qquad \text{when} \qquad \tau_s = \frac{L_s}{\lambda} \tag{4}$$

$$\frac{\lambda}{\mu N} < 1 \qquad \text{when} \qquad L_s = L_{q} + \rho \tag{5}$$

Here, a large-scale computing experiment are conducted, the data presented in Table 7 are used as the initial data and the numerical results shown in Table 8, Table 9 and Table 10 are obtained. [40-46]

Table 7 large-scale computing experiment ($m = 10 \; and \; \tau_s^* = 100$).

λ	μ	ρ
0.001	0.010	0.1
0.005	0.016	0.3
0.010	0.020	0.5
0.015	0.021	0.7
0.020	0.022	0.9

$L_q = f(\rho)$, $L_s = f(\rho)$ have been studied (Table 8), $L_q(L_s)$ maximum price $\rho = 0.3 * (0.5)$ prices are obtained, and the service process after these prices $L_q(L_s)$ is followed by a tendency to decrease prices.

Table 8 large-scale computing experiment.

ρ	L_q	L_s
0.1	1.05266315789504161E-11	0.14350070001052632
0.3	3.1047753832124043E-11	0.3145059003334771
0.5	2.104626871380261E-12	0.5000000002004625
0.7	1.556307260232261E-09	4.4962066983534367E-12
0.9	1.0295045078373537E-13	3.4918851425784085E-13

Table 9, $\tau_q = f(\lambda)$, $\tau_s = f(\lambda)$ dependencies were taken. Here λ With the increase in prices τ_s the price decreases and the optimal price of the system's characteristics $\lambda = 0.020$ when taken.

Table 9 optimal price ($\tau_s^* = 100$)

λ	τ_q	τ_s
0.001	1.052631578950416E-8	100.00000001052632
0.005	6.669550766424808E-8	62.500000066695506
0.010	2.004626871380261E-8	50.000000020046265
0.015	1.3042048401548407E-8	47.619047633208966
0.020	5.497522539186769E-9	45.45454546004298

It should be noted that with the increase in the price (Table 10), the total cost is considered minimum.

Table 10 optimal price ($m = 10\ and\ \tau_s^* = 100$)

N	P_1	P_2	P
2	9.003893535306852E-10	9.00389353404013E-10	1.8007787069346985-9
3	5.011567178659951E-10	5.011567178241353E-10	1.0023134356901306E-9
4	4.499506698714965E-10	4.4995066983534367E-10	8.99901337068401E-10
5	2.473885142689683E-10	2.4738851425784085E-10	4.94777028526809E-10

For the maximum price for the length of the queue in analytical model for limited and unlimited queue system.
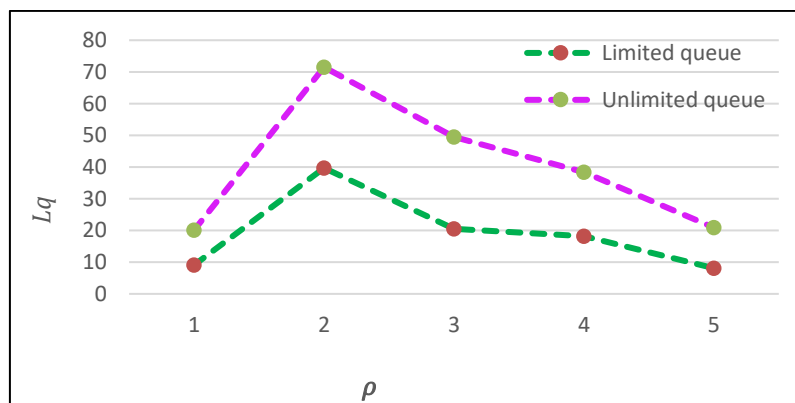
Fig. 4 Dependence to the process of limited and unlimited queue service.
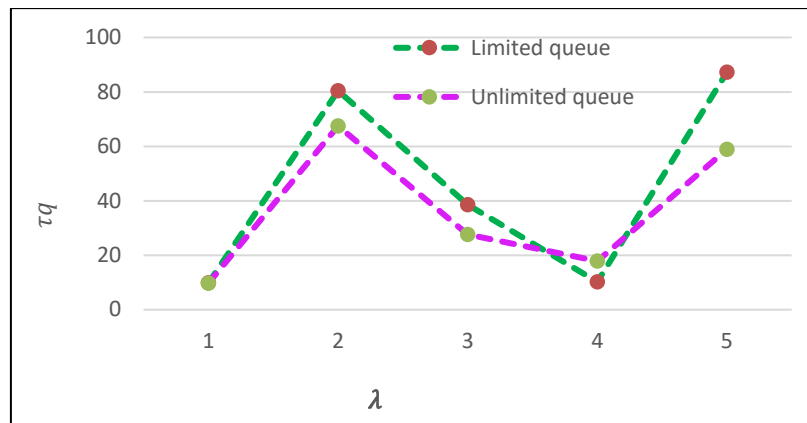
Fig. 5 Dependence to the process of limited and unlimited queue service.

In the first case, the length of queue is accompanied with maximum value and then the service process is accompanied with the downward trend in prices.

In this case, a rise occurs when there are optimal parameters of the system characteristics.

Based on the results obtained for the second option, the price of the length of the queue and the number of claims in the system are reduced.

A comparative analysis of the obtained results shows that these results differ from each other in a very short range and they can be used in building the distributed service networks of different types.

It should be noted that since it is difficult to analyze the different values of the input parameters of the system characteristics and analyze them against distribution laws, detailed analysis should be carried out by the simulation model.

It should be noted that the simulation model of queuing service systems in GPSS modeling language has been developed to verify the adequacy of analytical models, large-scale experiment has been conducted and results have been obtained. Based on the obtained results, the length of the queue of the five realizations of the exchangeable imitation model, the changing trend of the waiting time of the claims in the shift are as follows (**Fig. 6-9**).
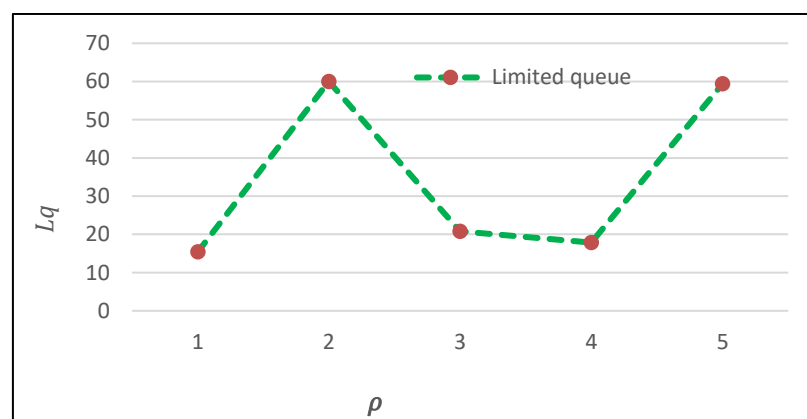
Fig. 6 The tendency towards changing the length of the queue according to the realizations of the imitation model for the limited queue service.
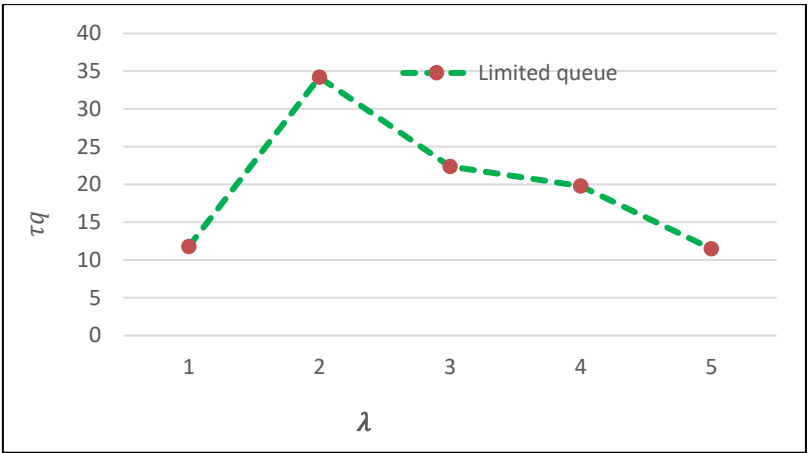
Fig. 7 The tendency towards changing the waiting time for the enquiries in the queue according to the realizations of the imitation model for the limited queue service.
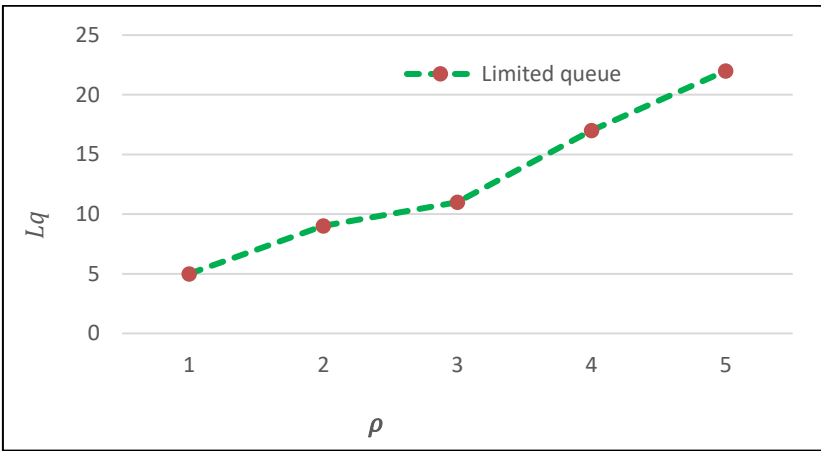
Fig. 8 The tendency towards changing the ratio to which the service equipment are used for the realizations of the imitation model for the process of the limited queue service.
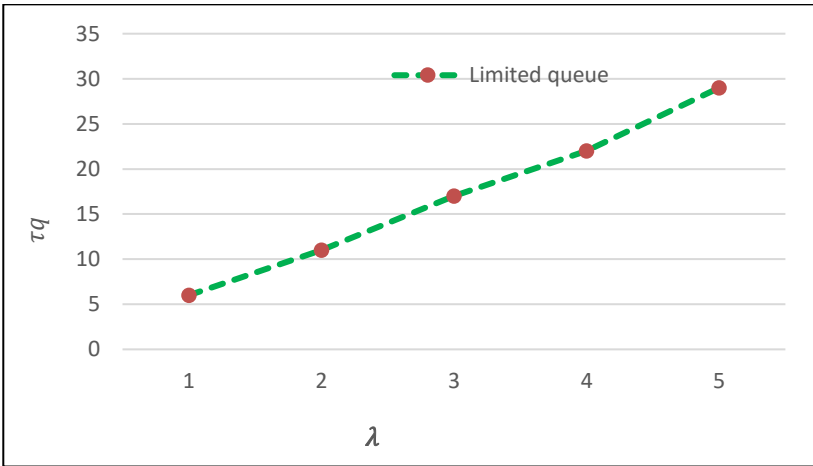
Fig. 9 The tendency towards changing the waiting time ratio for using the service equipment for the realizations of the imitation model for the process of the limited queue.

The length of queue and tendency of change in waiting time of request has been studied in both models. The time of acceptance of requests in the system is limited and exceeding this limit causes the loss of requests. Here, the issue of minimization of the loss of requests on the account of the removal of a part of requests from the system without waiting for the completion of service has been resolved as a

preventive measure and relevant effective "intellectual" algorithm has been developed. The developed algorithms of the organization of the interaction between the queue and service in QS model enables to determine the optimum value of the length of queue in terms of the real time scale of requests of system developers of some QS model for N computers and the procedure for the removal of "unpromising" requests. Along with the fact that the reviewed model covers a broad class of modern networks of different purpose, it confirmed that the forced removal of a certain part of requests waiting in the queue for this class is always advisable. Besides, this enables reducing the burden on service facilities, minimizing the loss of both types, and applying advanced software and technical support mechanisms. It also allows a reduction in total loss of requests. Therefore, the availability of parameters minimizing the total loss in the system for all options of interaction of software and technical support is certain. The expediency of such optimization does not depend on a specific situation and the certain complexity of the operating system is compensated with the increase in the productivity of the system. The tendency of change of the length of queue on the realizations of analytical and simulation models of limited and unlimited queue service processes is given in the Table 11 and 12.

Table 11 Analytical and simulation models of limited and unlimited queue service processes

Realizations	Analytical model L_Q	Simulation model L_Q
1	11.001	11.001
2	72.004	77.032
3	40.008	43.003
4	9.0050	10.100
5	86.012	84.000

Table 12 Analytical and simulation models of limited and unlimited queue service processes

Realizations	Analytical model L_Q	Simulation model L_Q
1	78.023	19.000
2	81.034	61.004
3	93.002	24.007
4	97.108	19.003
5	100.01	99.011

The tendency of change of waiting time of requests in the queue on the realizations of analytical and simulation models of limited and unlimited queue service processes is given in the tables 13 and 14.

Table 13 Simulation models of limited and unlimited queue service processes.

Realizations	Analytical model L_Q	Simulation model L_Q
1	5.000	11.023
2	7.004	35.034
3	14.007	23.002
4	19.003	21.108
5	25.011	9.0101

Table 14 Simulation models of limited and unlimited queue service processes.

Realizations	Analytical model L_Q	Simulation model L_Q
1	20.005	17.001
2	71.021	43.004
3	43.004	24.007
4	37.100	19.005
5	14.000	9.012

The tendency of the results of analytical (A) and simulation (I) models is determined by the Equation (6).

$$\Delta = \left[\frac{|i - A|}{A} \right] * 100\% \qquad (6)$$

Using this Equation, the tendency of the results of analytical (A) and simulation (I) models was (2-9) %. This confirms the consistency of analytical and simulation models with each other. The obtained results of analytical and simulation models can be used in the development of distributed service networks of different purposes. The results obtained based on the procedure, the algorithm for the calculation of the optimum values of parameters of dynamic priorities in such system and simulation algorithms are valuable information for the system developers that lead to the adequacy of analytical model.

5. Conclusion

The proposed DPSS method is simulated and evaluated using GPSS simulator and then it is compared with analytical (A) and simulation model. The evaluation results indicate the efficiency of the proposed method for the waiting time, length of the queue on realizations, length of the shift on the realization and limited and unlimited queue service processes. The proposed method simulation results showed that this method is capable of decreasing the waiting time more stably and improving length of the queue. The analysis of limited queuing service processes and the optimum organization of queue and service results in maximum effectiveness of the system in terms of removing requests that do not meet the condition of the value of presence of requests in the system on lack of exceeding the possible maximum value before the beginning of service from the system or during service time. The study of unlimited queuing service processes is of theoretical and practical significance in solving the criteria and restrictions issues used in the formulation of the problem, and it enables meeting expectations in the development of the systems of different purposes. The results of analytical and simulation models of experiments conducted on both models are compared, and it is identified that they differ in the allowable limit (2-9) %. Thus, the results obtained based on comparative analysis

of analytical and simulation models of limited and unlimited queuing service systems prove the adequacy of analytical model. In all cases, there is the finite value of the length of queue minimizing the total loss in the wide range of the rules for the selection of requests for the service and service time with Poisson input stream and exponential distribution law.

Compliance with ethical standards
Conflict of interest Fatholah Dadgar Arablu declare that he has no conflict of interest.
Ethical approval this paper does not contain any studies with human participants by any of the authors.

ORCID

Fatholah Dadgar Arablu 🆔 https://orcid.org/0000-0003-4242-1922

References

[1]   He, Q. C., & Chen, Y. J. (2018). Revenue-maximizing pricing and scheduling strategies in service systems with flexible customers. *Operations Research Letters*, *46*(1), 134-137.

[2]   Åkesson, M., & Edvardsson, B. (2018). Customer roles from a self-service system perspective. *International Journal of Quality and Service Sciences*, *10*(2), 196-210.

[3]   Zhang, W., Guo, J., Gu, F., & Gu, X. (2018). Coupling life cycle assessment and life cycle costing as an evaluation tool for developing product service system of high energy-consuming equipment. *Journal of Cleaner Production*, *183*, 1043-1053.

[4]   Albahri, O. S., Albahri, A. S., Mohammed, K. I., Zaidan, A. A., Zaidan, B. B., Hashim, M., & Salman, O. H. (2018). Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: Taxonomy, open challenges, motivation and recommendations. *Journal of medical systems*, *42*(5), 80.

[5]   Choi, B. D., Shin, B. C., Choi, K. B., Han, D. H., & Jang, J. S. (1998). Priority queue with two-state Markov-modulated arrivals. *IEE Proceedings-Communications*, *145*(3), 152-158.

[6]   Zaminkar, M., Sarkohaki, F., & Fotohi, R. A method based on encryption and node rating for securing the RPL protocol communications in the IoT ecosystem. International Journal of Communication Systems, e4693.

[7]   Faraji-Biregani, M., & Fotohi, R. (2020). Secure communication between UAVs using a method based on smart agents in unmanned aerial vehicles. The Journal of Supercomputing, 1-28.

[8]   Fotohi, R., Nazemi, E., & Aliee, F. S. (2020). An Agent-Based Self-Protective Method to Secure Communication between UAVs in Unmanned Aerial Vehicle Networks. Vehicular Communications, 100267.

[9]   Zaminkar, M., & Fotohi, R. (2020). SoS-RPL: Securing Internet of Things Against Sinkhole Attack Using RPL Protocol-Based Node Rating and Ranking Mechanism. WIRELESS PERSONAL COMMUNICATIONS.

[10]   Sarkohaki, F., Fotohi, R., & Ashrafian, V. (2020). An efficient routing protocol in mobile ad-hoc networks by using artificial immune system. arXiv preprint arXiv:2003.00869.

[11]   Mabodi, K., Yusefi, M., Zandiyan, S., Irankhah, L., & Fotohi, R. (2020). Multi-level trust-based intelligence schema for securing of internet of things (IoT) against security threats using cryptographic authentication. The Journal of Supercomputing, 1-26.

[12]   Seyedi, B., & Fotohi, R. (2020). NIASHPT: a novel intelligent agent-based strategy using hello packet table (HPT) function for trust Internet of Things. The Journal of Supercomputing, 1-24.

[13] Fotohi, R., & Bari, S. F. (2020). A novel countermeasure technique to protect WSN against denial-of-sleep attacks using firefly and Hopfield neural network (HNN) algorithms. The Journal of Supercomputing, 1-27.

[14] Fotohi, R. (2020). Securing of Unmanned Aerial Systems (UAS) against security threats using human immune system. Reliability Engineering & System Safety, 193, 106675.

[15] Fotohi, R., Firoozi Bari, S., & Yusefi, M. (2020). Securing wireless sensor networks against denial‐of‐sleep attacks using RSA cryptography algorithm and interlock protocol. International Journal of Communication Systems, 33(4), e4234.

[16] Jamali, S., Fotohi, R., & Analoui, M. (2018). An artificial immune system based method for defense against wormhole attack in mobile adhoc networks. Tabriz Journal of Electrical Engineering, 47(4), 1407-1419.

[17] Jamali, S., & Fotohi, R. (2017). DAWA: Defending against wormhole attack in MANETs by using fuzzy logic and artificial immune system. the Journal of Supercomputing, 73(12), 5173-5196.

[18] Fotohi, R., Heydari, R., & Jamali, S. (2016). A Hybrid routing method for mobile ad-hoc networks. Journal of Advances in Computer Research, 7(3), 93-103.

[19] Jamali, S., & Fotohi, R. (2016). Defending against wormhole attack in MANET using an artificial immune system. New Review of Information Networking, 21(2), 79-100.

[20] Fotohi, R., Ebazadeh, Y., & Geshlag, M. S. (2016). A new approach for improvement security against DoS attacks in vehicular ad-hoc network. International Journal of Advanced Computer Science and Applications, 7(7), 10-16.

[21] Fotohi, R., & Jamali, S. (2014). A comprehensive study on defence against wormhole attack methods in mobile Ad hoc networks. International journal of Computer Science & Network Solutions, 2, 37-56.

[22] Fotohi, R., Jamali, S., Sarkohaki, F., & Behzad, S. (2013). An Improvement over AODV routing protocol by limiting visited hop count. International Journal of Information Technology and Computer Science (IJITCS), 5(9), 87-93.

[23] Fotohi, R., Jamali, S., & Sarkohaki, F. (2013). Performance Evaluation of AODV, LHC-AODV, OLSR, UL-OLSR, DSDV Routing Protocols. International Journal of Information Technology and Computer Science (IJITCS), 5, 21.

[24] Fotohi, R., & Effatparvar, M. (2013). A cluster based job scheduling algorithm for grid computing. International Journal of Information Technology and Computer Science (IJITCS), 5(12), 70-77.

[25] Mandjes, M., Mannersalo, P., & Norros, I. (2005, January). Priority queues with Gaussian input: a path-space approach to loss and delay asymptotics. In Proceedings of the 19th International Teletraffic Congress (ITC-19) (pp. 1135-1144).

[26] Mannersalo, P., & Norros, I. (2001). Approximate formulae for Gaussian priority queues. In Teletraffic Science and Engineering (Vol. 4, pp. 991-1002). Elsevier.

[27] Mazzini, G., Rovatti, R., & Setti, G. (2005). A closed form solution of Bernoullian two-classes priority queue. IEEE Communications Letters, 9(3), 264-266.

[28] Barford, P., Bestavros, A., Bradley, A., & Crovella, M. (1999). Changes in web client access patterns: Characteristics and caching implications. World Wide Web, 2(1-2), 15-28.

[29] Leland, W. E., Taqqu, M. S., Willinger, W., & Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). IEEE/ACM Transactions on Networking (ToN), 2(1), 1-15.

[30] Liang, Q. (2002). Ad hoc wireless network traffic-self-similarity and forecasting. IEEE Communications Letters, 6(7), 297-299.

[31] Paxson, V., & Floyd, S. (1995). Wide area traffic: the failure of Poisson modeling. IEEE/ACM Transactions on Networking (ToN), 3(3), 226-244.

[32] Tickoo, O., & Sikdar, B. (2003). On the impact of IEEE 802.11 MAC on traffic characteristics. IEEE Journal on Selected

areas in communications, *21*(2), 189-203.

[33] Beran, J., Sherman, R., Taqqu, M. S., & Willinger, W. (1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on communications*, *43*(234), 1566-1579.

[34] Ashour, M., & Le-Ngoc, T. (2003, December). Priority queuing of long-range dependent traffic. In *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE* (Vol. 6, pp. 3025-3029). IEEE.

[35] Iacovoni, G., & Isopi, M. (2003). Performance analysis of a priority system fed by self-similar Gaussian traffic.

[36] Quan, Z., & Chung, J. M. (2003, May). Priority queueing analysis of self-similar in high-speed networks. In *Communications, 2003. ICC'03. IEEE International Conference on* (Vol. 3, pp. 1606-1610). IEEE.

[37] Kleinrock, L. (1976). Queueing systems, volume II: computer applications.

[38] Ozawa, T. (1990). Alternating service queues with mixed exhaustive and K-limited services. *Performance Evaluation*, *11*(3), 165-175.

[39] Lee, D. S., & Sengupta, B. (1993). Queueing analysis of a threshold based priority scheme for ATM networks. *IEEE/ACM Transactions on Networking (TON)*, *1*(6), 709-717.

[40] Tosirisuk, P., & Chandra, J. (1990). Multiple finite source queueing model with dynamic priority scheduling. *Naval Research Logistics (NRL)*, *37*(3), 365-381.

[41] Arablou, F. D., & Alakbar, A. (2015). Analysis of the Analytical and Simulations Results with the Network Model and Dynamic Priority of Limited Buffer. *Analysis*, *127*(13).

[42] Ismaylov BG, FA Dadgar Analysis of the characteristics of the networks with dynamic priorities and a limited buffer. Information tehnologiimodelirovaniya and management. Voronezh, №3 (81) 2013 s.292-296.

[43] Song, K. J., & Melikov, A. Z. (2008). BG Ismailov Optimal dynamic assignment of priorities in queuing systems with two types of applications and Riga.

[44] Chao, H. J., & Pekcan, I. H. (1994, May). Queue management with multiple delay and loss priorities for ATM switches. In *Communications, 1994. ICC'94, SUPERCOMM/ICC'94, Conference Record,'Serving Humanity Through Communications.'IEEE International Conference on* (pp. 1184-1189). IEEE.

[45] Ismayilov BG design a distributed network of service production facilities and gas treatment // M .: Gas promyshlennost.NTS. 2001, №1. s.29-35.

[46] Kellogg, D. (2006). Toward a post-academic science policy: Scientific communication and the collapse of the Mertonian norms.