Comparison of Four Conceptual Models of a Queuing System in Service Networks

Stoyan Poryazov*, Velin Andonov* and Emiliya Saranova*

* Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Block 8, 1113 Sofia, Bulgaria, e-mails: stoyan@math.bas.bg;
velin_andonov@math.bas.bg; emiliya@cc.bas.bg

Abstract: *Queuing systems are important type of components of almost all service networks. Despite this, in the scientific literature, the graphical representation of their conceptual models is too simplified and one-sided. On the other hand, models of queues using Generalized nets are not studied enough. In the present paper, four different conceptual models of a queuing system, containing buffer and server, are described and compared. Three of the models are independent of the used analytical models while one of them is oriented towards the Generalized nets. Five problems connected to the graphical representation of queuing systems are formulated. The results can be used for illustration for the purpose of teaching, for the selection of analytical approach to the modeling of specific queuing systems and for the development of systems for graphical representation.*

Сравнение на четири концептуални модела на опашкови системи в обслужващи мрежи (Стоян Порязов, Велин Андонов, Емилия Саранова). *Опашковите системи са съществен тип компоненти на почти всички обслужващи мрежи. Въпреки това, в обширната научна литература, техните концептуални модели са представени графично по твърде опростен и едностранчив начин. От друга страна, моделите на опашки, използващи методите на обобщените мрежи, не са представени достатъчно. В настоящата статия са описани и сравнени четири различни концептуални модела на една опашкова система, съдържаща буфер и сървър, като три от тях са независими от използваните аналитични методи, а четвъртият е ориентиран към обобщените мрежи. Формулирани са пет проблема, свързани с графичното представяне на опашковите системи. Резултатите могат да са полезни при онагледяване за целите на преподаването, при избора на аналитичен подход за моделиране на конкретни опашкови системи и при разработване на системи за графично представяне.*

Introduction

For the queuing systems containing buffer and server, the first quantitative models are created by Erlang [1]. After that newer methods have been used [2]. These systems are used in practically all service networks, including telecommunication, computing, logistic, etc [3]. Despite this, in the broad scientific literature their conceptual models are being represented graphically in a very simple and one-sided way [4]. Models of queues, using the methods of the Generalized nets are not well studied. In the second section of the present paper basic monofunctional devices, their graphical representations and the used notation of parameters are described. In the third section, classical representation of the queuing systems is described in which the introduced notation is used. In the fourth section, an extension of the classical representation is described and the probability of blocking of the server is explicitly shown. In the fifth section, a queue is presented in detail with the use of five basic virtual devices. In the sixth section, a Generalized net model of the of the classical representation is described. As a conclusion, the four conceptual models of a queuing system are compared and recommendations for their use are made.

Basic virtual devices

At the bottom of the structural model presentation, we consider basic virtual devices that do not contain any other virtual devices. A basic virtual device has the graphic representation as shown in Fig. 1.

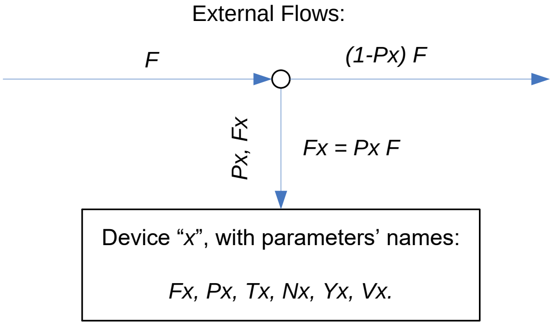

Fig. 1. A graphical representation of a basic virtual device x.

The parameters of the basic virtual device x are the following (c.f. [5] for terms definition):

Fx – Intensity or incoming rate (frequency) of the flow of requests (i.e. the number of requests per time unit) to device x;

Px – Probability of directing the requests towards device x;

Tx – Service time (duration of servicing of a request) in device x;

Yx – Traffic intensity [Erlang];

Vx – Traffic volume [Erlang - time unit];

Nx – Number of lines (service resources, positions, capacity) of device x.

In our models, we consider monofunctional idealized basic virtual devices of the following types (Fig. 2):

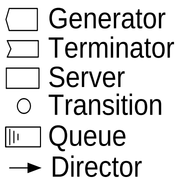

Fig. 2. Base virtual conceptual devices.

- Generator – this device generates calls (service requests, transactions);
- Terminator – this block eliminates every request entered (so it leaves the model without any traces);
- Director – this device unconditionally points to the next device, which the request shall enter, but without transferring or delaying it;
- Server – this device models the delay (service time, holding time) of requests in the corresponding device without their generation or elimination. It models also traffic and time characteristics of the requests processing (c.f. Fig. 3);
- Transition – this device selects one of its possible exits for each request entered, thus determining the next device where this request shall go to;
- Queue – this Buffer device contains a queue. In this paper the queuing discipline (e.g. FIFO, LIFO etc.) is not considered.

Classical representation of a queuing system

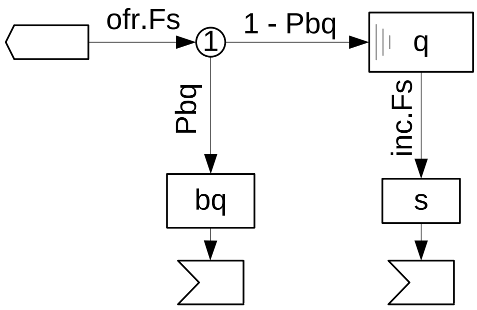

Fig. 3. Classical representation of a queuing system.

The Generator in Fig. 3 generates offered flow of call attempts, with frequency *ofr.Fs*, to the Server (s). If the server is not full, requests are incoming, with frequency *inc.Fs*, and are served in it. After this requests leave the model at the Terminator device after the *s* device. If the server is full, the offered requests wait in the Queue device (*q*). If the queue is full, the requests are transferred by the Transition 1, with probability *Pbq*, to the virtual device "blocked queuing" (*bq*), and leave the model in the Terminator device below it (see Fig. 3). The "blocked queuing" device is outside the queuing system. It corresponds to the duration of specific signalization, e.g. listening of the busy tone in telephone systems.

Extended classical representation

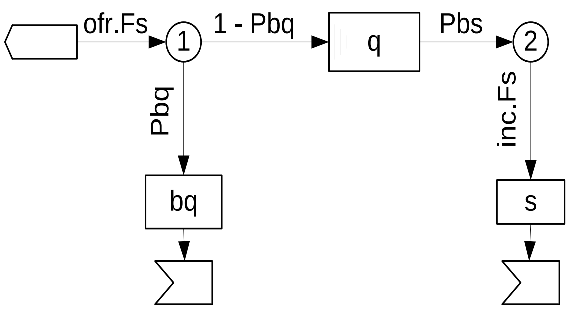

Fig. 4. Extended classical representation of a queuing system.

In Fig.4, in comparison with Fig. 3, Transition 2 is added. This allows explicit presentation of the probability *Pbs* of finding the Server full by the offered requests. With this probability requests are stopped by Transition 2 and wait, in the Queue (the *q* device), for free places in the Server (*s*).

This presentational approach requires the introduction of a new function of the transition blocks, apart from transferring – stopping the call attempts.

Detailed representation

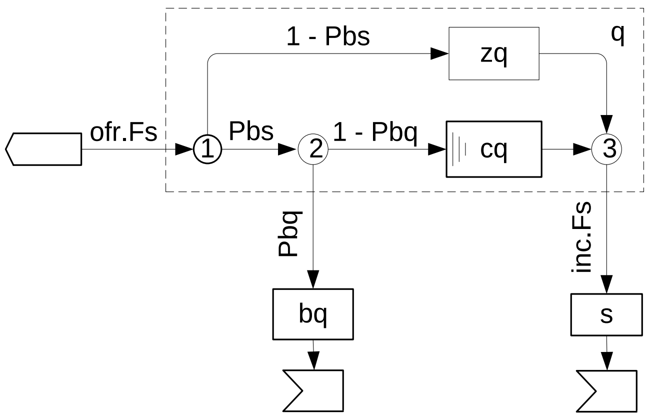

Fig. 5. More detailed representation of a queuing system.

In Fig. 5, an internal structure of the Queue is presented, including two virtual devices: "carried queue" (*cq*) and "zero queuing" (*zq*). Requests pass the Queue without delay, with probability *Pzq = 1-Pbs*, in case there are free places available in the Server, in the moment of their arrival. The duration of the zero queuing (*Tzq*) may be zero, or close to it. The total queuing time (*Tq*) is:

(1) Tq = Pbs (1-Pbq) Tcq + (1-Pbs) Tzq .

The approach used in Fig. 5 is a detailization of the Queue device in the classical approach (Fig. 3). It presents explicitly the important concept of zero queuing and the probability of blocked server (*Pbs*). It is more complex, but allows more clear and full presentation of the processes in the queuing system.

Generalized net model of a queuing system

Another approach to the conceptual modeling of a queuing system is to use the apparatus of the Generalized nets (see [6], [7]). Conceptual models of a part of overall telecommunication system using ordinary Generalized nets and Generalized nets with characteristics of the places (see [8]) are described in [9]. The use of Generalized nets as a tool for modeling of telecommunication systems has continued in [10] where a Generalized net model of a part of the Switching stage of overall telecommunication system with a queue is described.

Here, we propose a Generalized net model of a queuing system as an alternative approach to the conceptual model in Fig. 3. The graphical representation of the net is shown in Fig. 6. In order to present a more detailed description of the net, we consider FIFO queuing discipline.

The places of the net correspond to virtual devices in the following way:

- l_1 and l_2 represent the Generator before Transition 1 in Fig. 3;
- l_3 has no analogue in Fig. 3;
- l_4 and l_5 represent the "blocked queuing" (*bq*) device in Fig. 3;
- l_6 represents the Terminator device after the *bq* device in Fig. 3;
- $l_{q,1}$, $l_{q,2}$, ..., $l_{q,Nq}$, l_7 represent the Queue device in Fig. 3;
- l_8 and l_9 represent the *s* device (comprising device of the Switching system);
- l_{10} corresponds to the Terminator device after the *s* device in Fig. 3.

Each of the six transitions has the following meaning:

- Z_1 represents the function of the Generator before Transition 1 in Fig. 3;
- Z_2 represents the function of Transition 1 in Fig. 3;
- Z_3 represents the function of the Director between the *bq* device and the Terminator device in Fig. 3;
- Z_4 represents the service of the call attempts in the Queue;
- Z_5 represents the service of call attempts in the Switching system (the *s* device in Fig. 3);
- Z_6 represents the function of the Director between the Switching system and the Terminator in Fig. 3.

Four different types of tokens are used in the model.

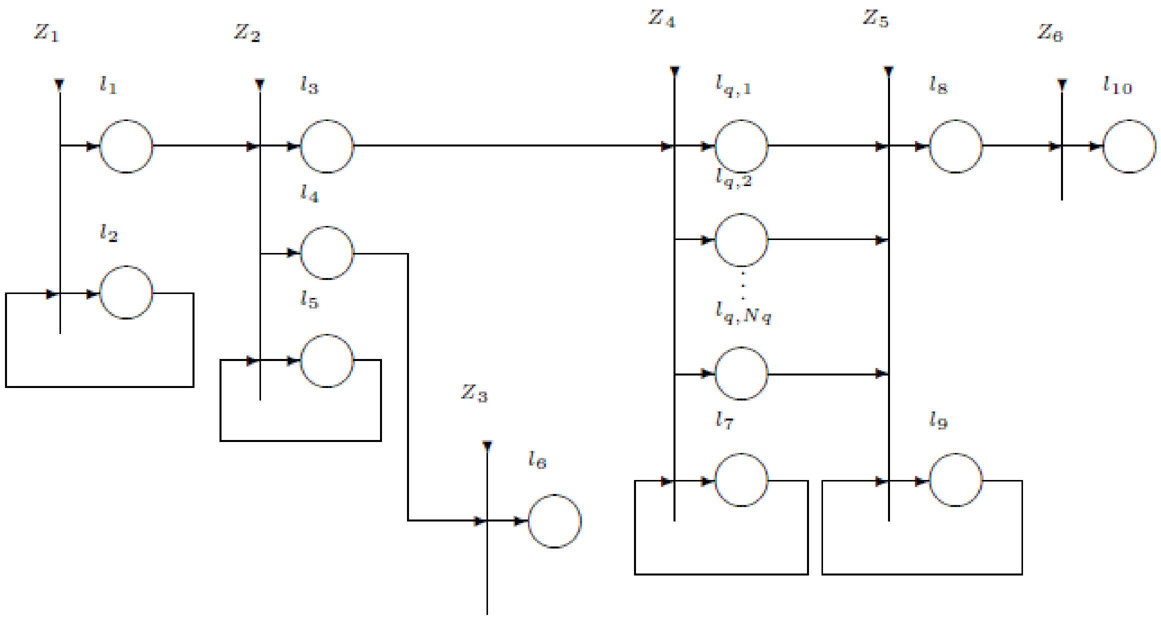

Fig. 6. Generalized net model of a queuing system.

- Tokens of type α represent the call attempts. In the initial moment token of type α stays in place l_2 with characteristic "formula for generating the flow of call attempts". During the functioning of the net, a token of type α enters place l_1 coming from l_2 with characteristic "volume, duration of service,call destination";
- Token of type β stays in place l_5 in the initial time moment. It is used to accumulate data about the *bq* device;
- Token of type γ stays in place l_7 in the initial moment. It is used to accumulate data about the *q* device;
- Token of type δ stays in place l_9 in the initial moment. It is used to accumulate data about the *s* device.

All tokens except the tokens of type α have initial characteristic: "initial values of Y_{dn}; P_{dn}; F_{dn}; T_{dn}" where the subscript *dn* should be replaced by the corresponding device name.

We present detailed description of each of the transitions of the Generalized net. The Generalized net which is used is a reduced one (see [7]) and every transition has the form *Z* = <*L'*, *L''*, *r*> where *L'* denotes the set of input places of the transition, *L''* denotes the set of output places of the transition and *r* is the index matrix of the transition's condition.

$$Z_1 = \langle \{l_2\}, \{l_1, l_2\}, r_1 \rangle ,$$

where

$r_1 =$

	l_1	l_2
l_2	$W_{2,1}$	true

and $W_{2,1}$ = "A flow of call attempts is generated". When the truth value of predicate $W_{2,1}$ is "true" the α token in place l_2 splits into two tokens – the same α token which continues to stay in place l_2 and token α_1 which enters place l_1 with characteristic "volume, duration of service, call destination".

$$Z_2 = \langle \{l_1, l_5\}, \{l_3, l_4, l_5\}, r_2 \rangle ,$$

where

$r_2 =$

	l_3	l_4	l_5
l_1	$W_{1,3}$	$W_{1,4}$	$W_{1,5}$
l_5	false	false	true

and

- $W_{1,3}$ = "$Y_q < N_q$".
- $W_{1,4} = W_{1,5} = \neg W_{1,3}$.

When the truth value of the predicate $W_{1,3}$ is "true" the token of type α enters place l_3 without obtaining any new characteristic. When the truth value of the predicate $W_{1;4}$ is "true" the token of type α in place l_1 splits into two identical tokens one of which enters place l_4 and the other one merges with the β token in place l_5. In place l_4 the tokens do not obtain new characteristics. Token β in place l_5 obtains the characteristic "current value of Y_{bq}".

$$Z_3 = \langle \{l_4\}, \{l_6\}, r_3 \rangle,$$

where

$r_3 =$

	l_6
l_4	true

In place l_6 the tokens of type α leave the net.

$$Z_4 = \langle \{l_3, l_7\}, \{l_{q,1}, l_{q,2}, \ldots, l_{q,Nq}, l_7\}, r_4 \rangle,$$

where

$r_4 =$

	$l_{q,1}$	$l_{q,2}$	…	$l_{q,Nq}$	l_7
l_3	$W_{3,q1}$	$W_{3,q2}$	…	$W_{3,qNq}$	$W_{3,7}$
l_7	false	false	…	false	true

and

- $W_{3,qi}$ = "The output $l_{q,i}$ is the highest priority empty place among the places $l_{q,1}; l_{q,2}; \ldots; l_{q,Nq}$." for $i = 1, 2, \ldots, N_q$;
- $W_{3,7} = W_{3,q1} \lor W_{3,q2} \lor \ldots \lor W_{3,qNq}$.

In order to employ FIFO queuing discipline, each of the output places $l_{q,1}$, $l_{q,2}$, …, $l_{q,Nq}$ should have capacity 1 and their priorities can be such that $\pi_L(l_{q,1}) > \pi_L(l_{q,2}) > \ldots > \pi_L(l_{q,Nq})$. When one of the predicates $W_{3,qi}$ for $i = 1, 2, \ldots, N_q$ has truth value "true" the α token in place l_3 splits into two identical tokens the first of which enters the corresponding output place among places $l_{q,1}$, $l_{q,2}$, …, $l_{q,Nq}$ without obtaining new characteristic. The other one enters place l_7 where it merges with the γ token. Token γ obtains the characteristic "current value of Y_q; T_q; F_q, list of all tokens in the output places $l_{q,1}$, $l_{q,2}$, …, $l_{q,Nq}$ and the duration of their stay in the place".

$$Z_5 = \langle \{l_{q,1}, l_{q,2}, \ldots, l_{q,Nq}, l_9\}, \{l_8, l_9\}, r_5 \rangle,$$

where

$r_5 =$

	l_8	l_9
$l_{q,1}$	$W_{q1,8}$	$W_{q1,9}$
$l_{q,2}$	$W_{q2,8}$	$W_{q2,9}$
⋮	⋮	⋮
$l_{q,Nq}$	$W_{qNq,8}$	$W_{qNq,9}$
l_9	false	true

and

- $W_{qi,8}$ = "The current token in place $l_{q,i}$ has stayed more time in the place than the tokens in all places $l_{q,1}$, $l_{q,2}$, …, $l_{q,Nq}$ and $Y_s < N_s$", for $i = 1, 2, \ldots, N_q$;
- $W_{qi,9} = W_{qi,8}$, for $i = 1, 2, \ldots, N_q$.

When one of the predicates $W_{qi,8}$ for $i = 1, 2, \ldots, N_q$ has truth value "true" the α token in the corresponding input place splits into two identical tokens the first of which enters place l_8 without obtaining new characteristic. The other one enters place l_9 where it merges with the δ token. Token δ obtains the characteristic "current value of Ys; Ts; Fs".

$$Z_6 = \langle \{l_8\}, \{l_{10}\}, r_6 \rangle,$$

where

$r_6 =$

	l_{10}
l_8	true

.

In place l_{10} the tokens of type α leave the net.

Comparison of the models

One problem in the presented models is that different basic functions of virtual devices with the same graphic representation are used. For instance, Transition 1 in Fig. 3 distributes the call attempts of the incoming flow into two outgoing. Transition 2 in Fig. 4 holds the call attempts without changing their path. Transition 3 in Fig. 5 joins the incoming call attempts from different paths in one outgoing path. The transitions in Fig. 6 have different graphical representation and also have four different functions. These functions are: (i) synchronization of the incoming call attempts from all incoming paths; (ii) terminating all incoming call attempts; (iii) generating new outgoing call attempts; (iv) synchronization of the departure of the newly generated call attempts. Similar functions have the transitions of the Petri Nets [11].

Second problem is the selection of a functional normalization of the use of base virtual devices. It is desirable, for the purpose of education and research, the different functions to be represented graphically in different ways in order to avoid confusion.

Third problem is the representation of different types of servicing in the queue. Only in Fig. 5 the service without waiting is represented graphically in spite of it being included in all serious analytical models. For educational purpose, the visualization of the servicing without waiting is preferable while for experienced analysers it unnecessarily complicates the graphics.

Fourth problem is the visual representation of generalizations of the service operations. For instance, the block q in Fig. 5 is a detailization of the the block q in Fig. 3. The representation in Fig. 4 is not a very successful one because it is not clear why the call attempts are directed with probability *Pbs* to Transition 2 while in reality they are hold in the buffer. The term Queuing System is a conceptual generalization of the devices buffer and server (see Fig. 3). The possibilities for generalization and detailization are very important for the education, the projection of new and the maintenance of existing service systems.

Fifth problem is to what degree the graphical representation helps for the derivation of analytical models. It may turn out that the state-transition diagram is the most suitable representation for the derivation of the needed formulae. It is, however, too detailed to be included in a graphical model of the overall network.

Conclusions

On the basis of the presented different conceptual models of one queuing system and their analysis the following conclusions can be made:

- Different graphical representations of one queuing system are possible.
- The representations should take into consideration the level of the users' preparation and the problem that is to be solved.

Problems for future research include the development of a base system of monofunctional virtual devices which can be used for the complete representation of all service networks, allow graphical generalization and are suitable for the derivation of analytical models.

Acknowledgements

The work was partially funded by the Bulgarian NSF Projects DM 12/2 – "New Models of Overall Telecommunication Networks with Quality of Service Guarantees".

REFERENCES

[1] Erlang, A. Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. Elektrotkeknikeren, 13: 5, 1917.

[2] Kleinrock, L. Queueing Systems, Volume 1: Theory, 1975.

[3] Otsetova, A. Validation of the Logistics Service Quality Scale in Bulgarian Courier Sector. Management and Education, Vol. XII, (1) 2016, 46-52.

[4] Shaguna Gupta, Sakshi Arora. Queuing Systems in Cloud Services Management: A Survey. International Journal of Pure and Applied Mathematics, Volume 119, No. 12, 2018, 12741-12753 (url: http://www.ijpam)

[5] ITU-T Recommendation E.600 (03/93), Terms and definitions of traffic engineering.

[6] Atanassov, K. Applications of Generalized Nets. World Scientific Publ. Co., Singapore, 1993.

[7] Atanassov, K. On Generalized Nets Theory. Prof. M. Drinov Academic Publ. House, Sofia, 2007.

[8] Andonov, V., Atanassov, K. Generalized nets with characteristics of the places. Compt. Rend. Acad. Bulg. Sci. 66(12), 2013, 1673-1680.

[9] Poryazov, S., Andonov, V., Saranova, E. Comparison of Conceptual Models of Overall Telecommunication Systems with QoS Guarantees. In: Christiansen H., Jaudoin H., Chountas P., Andreasen T., Legind Larsen H. (eds) Flexible Query Answering Systems. FQAS 2017. Lecture Notes in Computer Science, vol 10333. Springer, Cham, 2017.

[10] Andonov, V., Poryazov, S., Otsetova, A., Saranova, E. Generalized net model of a part of overall telecommunication system with a queue in the Switching stage. Proc. of the 16th International Workshop on Generalized Nets, Sofia, 9-10 February 2018 (in press).

[11] Petri, C.A. Kommunikation mit Automaten. Institut für Instrumentelle Mathematik der Universität Bonn, Schriften des IIM Nr. 2 (1962).