ITC 11, Minoru Akiyama (Editor)
Elsevier Science Publishers B. V. (North-Holland)

TRANSIT DELAY DISTRIBUTIONS IN PRIORITY QUEUING NETWORKS

Paul J. KUEHN and Wolfgang SCHMITT

Institute of Communications Switching and Data Technics
University of Stuttgart, Fed.Rep.of Germany

Department of Communications
University of Siegen, Fed.Rep.of Germany

ABSTRACT

The paper presents an analysis procedure for transit delay distributions of closed queuing networks with preemptive or nonpreemptive priorities. The procedure is developed in two steps: at first, the decomposition of the priority queuing network with P classes of priorities into P queuing networks, one for each priority class, and, secondly, the analysis of the transit delay distribution function for a single-priority class network. The first step is based on a decomposition technique for priority stations into load- and queue length-equivalent single-class stations with state-dependent service rates. In the second step, such networks are analyzed by a first passage time method exactly as well as approximately by network transformations. The procedure involves several approximative assumptions which are validated by computer simulations.

1. INTRODUCTION

1.1 Problem

Transit delays are of primary interest for applications of interactive computer systems (terminal-I/O), paged computer systems (disk-I/O), computer communications networks (packet delays), or signalling networks based on the common channel interoffice signalling system (call setup delays). Modelling of such cases leads to closed or open queuing networks.

Queuing networks have received, therefore, much attention in research during recent years. Most research has been concentrated on average delays of product-form queuing networks [1,2], decomposition and aggregation techniques [3,4,5] in case of networks with a complex structure or general service centers. More recently, first approaches have been made towards multi-class queuing networks with priorities [6,7,8,9,10].

Transit delay distributions have been analyzed only recently in case of closed Markovian queuing networks [11,12,13,14,15]. The distribution of transit delays allows a much deeper insight into the network's behaviour as, e.g., for the analysis of the percentiles in case of terminal response times, post-dialling call delays, dial-tone delays, packet-acknowledgement delays, etc. This analysis, however, is more difficult and requires time-dependent processes even in the stationary case.

Models of the real world of computer systems and communications switching control lead often to priority queuing networks which have to be analyzed with respect to the average delays and with respect to the distribution of transit delays such as response times, cycle times, or flow times.

An example of the class of models treated in this paper is shown in Fig. 1.

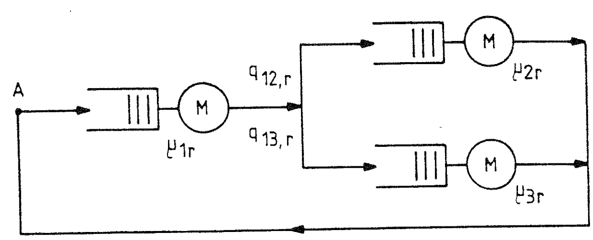

Fig. 1. Multi-Class Priority Queuing Network

The central server type model of Fig. 1 consists of 3 Markovian service stations. There are P classes of customers with populations M_r for class r, r = 1,2,..., P. The service rates μ_{ir} depend on the station number i and on the priority class r. Each class of customers is routed independently according to a class-individual probabilistic routing matrice $Q_r = (q_{ij,r})$, where $q_{ij,r}$ denotes the probability of a class-r customer being routed to station j after leaving station i. Within each station or, at least within one of the stations, customers are scheduled for service according to a preemptive or nonpreemptive priority discipline. The queue discipline within each station is FIFO for each class; other disciplines could be principally included, as well.

We are interested in the transit delay distribution of a priority r-test customer, i.e. the distribution of the time elapsing betweeen two successive passages of that r-test customer through the control point A, also called the cycle time of a class r-customer. Similar problems arise in the analysis of the transit delay for the passages of an r-test customer between two arbitrary points within the network.

1.2 Outline of the Analysis

The analysis is based on some recent results for both, the transit delay analysis for single-class networks and decomposition techniques for multi-class priority networks. The two solution steps (1) and (2) are outlined in Chapter 2 and 3:

11 th International Teletraffic Congress (ITC), Kyoto/Japan, September 4 - 9, 1985.

(1) Decomposition of the P-class network into P single-class networks for each of the priority classes. This decomposition provides equivalent service centers for each considered priority class. The equivalent service centers own a state-dependent service rate which reflects the influence of all other priority class customers. This approximate decomposition method has been validated for average queuing delays.

(2) Transit delay analysis of the equivalent class r-network. The transit delay distribution analysis rests on the method of first passage times. The fate of a class r-customer is explicitly considered and described by backward-type differential equations. This method allows also for state-dependent service rates.

Note that the method of step 2 is principally also applicable to solve the given problem of transfer delays within multipriority networks. The state description technique, however, requires multiple variables per station so that the resulting systems of differential equations are too complex to solve.

In case of more complex network structures, both steps (1) and (2) can be combined with aggregation methods to replace a considered station's complete environment by one composite server.

In the paper we develop both steps (1) and (2) separately. Approximate methods will be validated by simulation results.

The combination of both steps to the transit delay analysis of priority class queuing networks will be exemplified by some network examples in Chapter 4. Since the procedure involves several approximative assumptions, the analytical results are validated by computer simulations.

2. DECOMPOSITION OF PRIORITY QUEUING NETWORKS

2.1 Decomposition Principle

The method we will apply to decompose the priority stations of a queuing network with two priority classes is described comprehensively in [8] and [10]. Hence, concerning this method, we confine ourselves to a summary of the essentials.

The decomposition of a two-class priority queuing model into two single-class queuing models without a priority discipline is best explained by reference to Fig. 2, where:

λ_r	arrival rate
μ_r	service rate
x_r	number of class-r customers in the system
$r = 1,2$	class index; r = 1: high priority class r = 2: low priority class
$\mathbb{N}$	set of positive integers
$:=$	equals by definition

We call these single-class queuing models "Virtual Server" models. This is due to the fact that each class is assumed to receive its service from a dedicated Virtual Server, whose service time distribution function is chosen such, that the priority mechanism is appropriately taken into account. Replacing the original server by these

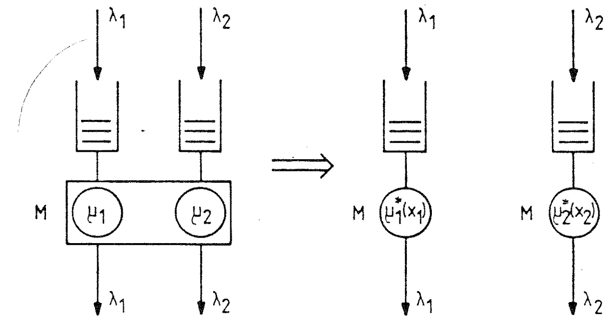

Fig. 2. Decomposition of a priority queue into Virtual Server models with state-dependent reduced service rates

Virtual Servers, a two-class priority queuing network is transformed into a queuing network without a priority discipline.

An important feature of the Virtual Server models presented in Fig. 2 are the state-dependent reduced service rates $\mu_r^*(x_r)$ (r=1,2) defined by Eqs. (2.1a-b):

$$\mu_r^*(x_r) := c_r(x_r)\cdot\mu_r, \qquad (2.1a)$$

$$c_r(x_r) := \text{prob(class r-customer in service} \mid x_r \text{ class r-customers in system)}$$

$$x_r \in \mathbb{N};\ r=1,2\ . \qquad (2.1b)$$

It has been proved in [9] that this choice of the service rates provides an exact description of the marginal distributions $p_r(x_r)$ (r=1,2) of system states in the priority queuing models M/M/1/PRE and M/M/1/NONPRE. Concerning the global behaviour of traffic flow, our approach is still an approximation. Particularly, the output streams are in general not Poisson processes.

We call our decomposition method the state-dependent reduced occupancy approximation and refer to it as to the s d r o a .

2.2 Test-Bed Examples for Decomposition

In this section we illustrate the application of our decomposition method by means of two test-bed queuing networks.

First, we consider the cyclic queuing model •/M/1/PRE → •/M/1/PRE (see Fig. 3a). Decomposition of the priority stations leads to two single-class queuing models depicted in Fig. 3b. In accordance with Eqs. (2.1a-b), the service rates $\mu_{ir}^*(x_{ir})$ (i=1,2; r=1,2) are determined by the following equations:

$$\mu_{i1}^* = \mu_{i1} \quad ,\ i=1,2 \qquad (2.2a)$$

$$\mu_{i2}^*(x_{i2}) = c_{i2}(x_{i2})\cdot\mu_{i2} \quad ,\ i=1,2 \qquad (2.2b)$$

where

$$c_{i2}(x_{i2}) = \text{prob}(X_{i1}=0 \mid X_{i2}=x_{i2}),\ x_{i2}=0,1,\ldots,M_2;\quad i=1,2. \qquad (2.2c)$$

The random variable X_{ir} denotes the number of class-r customers at station i and M_r denotes the network population of class r-customers.

In the present case the calculation of the unknown reduction coefficients $c_{i2}(x_{i2})$ can be carried out exactly by setting up the global balance equations and solving them (see [7]). For instance let

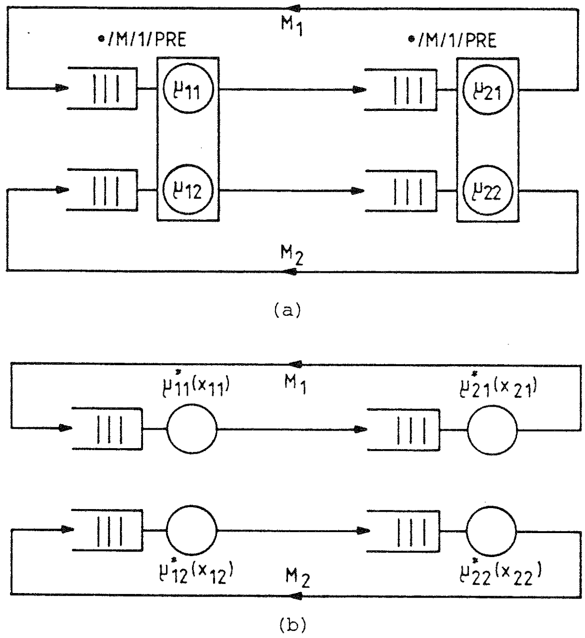

Fig. 3. (a) Cyclic queuing model •/M/1/PRE → •/M/1/PRE

(b) Decomposition into two single-priority networks

$M_1 = M_2 = 2$

$\mu_{11} = 0.2, \quad \mu_{12} = 1.0,$

$\mu_{21} = 0.1, \quad \mu_{22} = 0.1.$

Then we get

$c_{12}(1) = 0.11475, \quad c_{12}(2) = 0.07018,$

$c_{22}(1) = 0.50820, \quad c_{22}(2) = 0.09091.$

It must be mentioned here that the exact values for the coefficients $c_{i2}(x_{i2})$ lead to an exact description of the marginal distribution $p_{i2}(x_{i2})$ of the cyclic queuing model •/M/1/PRE → •/M/1/PRE. For a proof see [10].

Unfortunately, an exact calculation of the reduction coefficients is only possible for priority networks of little complexity. If this requirement will not be fulfilled, we rely on appropriate approximation techniques (see [10]). In order to distinguish this approach from the s d r o a we call it the modified s d r o a and refer to it as to the m s d r o a .

We have applied such an approximation technique to decompose our second test-bed queuing network for preemptive priorities with a structure as depicted in Fig. 1. The procedure to calculate the reduction coefficients for this network type is described in section 3.2 of [10]. Here, we give only some numerical values for the reduction coefficients $c_{i2}(x_{i2})$.

Let $M_1 = M_2 = 2$, $\mu_{ij} = 1.0$ ($i=1\ldots3$; $r=1,2$),

$q_{12,r} = q_{13,r} = 0.5$ ($r=1,2$).

Then, $c_{12}(1) = 0.25609 \qquad c_{12}(2) = 0.18085$

$c_{22}(1) = 0.51291 \qquad c_{22}(2) = 0.42324$

$c_{32}(1) = 0.51291 \qquad c_{32}(2) = 0.42324$

Similar results have been obtained for nonpreemptive priorities, see [9]. The decomposition principle has been validated by computer simulations for a wide parameter range which has shown an acceptable accuracy.

3. TRANSIT DELAY ANALYSIS IN SINGLE-CLASS MARKOVIAN NETWORKS

3.1 Transit Delay as First Passage Time Problem

We consider a class of Markovian queuing models which can be discribed by a Markov chain with an enumerable set of states and a continuous time parameter. The behavior of the Markov chain can be described by the well-known Chapman-Kolmogorov relation from which two sets of differential equations for the transition probabilities can be derived, the Kolmogorov forward equations and backward equations [16].

The waiting time or the response time of a customer can be considered as "life times" T of one or several tagged (test) customers within a properly defined set S of states. The life time terminates when the test customer leaves S for the first time entering a "taboo" set $H = \overline{S}$; his life time is equal to the "first passage time" to H. The life time process can be considered as a special process with "absorbing" states in H. This modified process can be constructed from the system state transition probabilities under the condition that states in H are excluded. The state of the modified process must be specified such that all effects which may influence the life time T of the test customer directly or indirectly are reflected properly.

The general procedure of this life time process analysis has been outlined in [15]. Here, we refer only to the main results.

Let $w(t|i)$ denote the conditional complementary life time distribution function (df) where the considered life time of a test customer has started at initial state i. Then, the life time process is described by the set of Kolmogorov backward-type equations

$$\frac{d}{dt} w(t|i) = -q_i w(t|i) + \sum_{j \neq i} q_{ij} w(t|j) \qquad (3.1a)$$

where

$$q_i = \sum_{j \neq i} q_{ij} + \varepsilon_i \quad , \; i,j \in S. \qquad (3.1b)$$

Within Eqs. (3.1a,b) q_{ij} denotes the instantaneous rate of transitions from state i to state j in S, q_i the rate for transitions from i to any other state in S or H, and ε_i the life time terminating rate from state i into H. These rates are found by considering the underlying life time process.

From Eqs. (3.1a), a set of linear equations can be derived for the ordinary k-th moments of the conditional life time $m_i^{(k)}$:

$$q_i m_i^{(k)} - \sum_{j \neq i} q_{ij} m_j^{(k)} = k m_i^{(k-1)} \quad , \qquad (3.2)$$

where $m_i^{(0)} = 1, \quad i,j \in S,\; k \in \mathbb{N}.$

The total complementary life time df is composed from the initial state distribution $\Pi(i)$ and the conditional life time df's $w(t|i)$ according to

$$W(t) = \sum_{i \in S} \Pi(i) \cdot w(t|i) \qquad (3.3)$$

3.2 Exact Analysis

The exact analysis of the life time process will be demonstrated for the example of a cycle time analysis of a closed queuing network with two stations, as shown in Fig. 3b, where the service rate $\mu_i(x_i)$ of station i depends on the number of customers x_i in that station, i = 1,2. The queue discipline in each station is FIFO.

The exact result of the cycle time df for this model is known only in the simpler case of constant service rates, see [11,13,15].

Now we consider the case where the instantaneous service rate of a server may depend on the actual number of customers in that station. Combinatorial methods for the cycle time analysis are not adequate since in this case succeeding customers in a queue behind the test customer may influence his cycle time even if overtaking is not possible. The state description has now to be augmented.

Let $\zeta(t) = (i,j,k)$ define the state of cycle time process, where

k indicates the station where the test customer is currently located in

i = number of customers in station 1 consisting of the test customer and his predecessors in line (k = 1)
or
total number of customers in station 1 (k = 2)

j = total number of customers in station 2 (k = 1)
or
number of customers in station 2 consisting of the test customer and his predecessors in line (k = 2).

The state transition diagram is shown in Fig. 4 for the special case of M = 3 (extension to general M is straightforward).

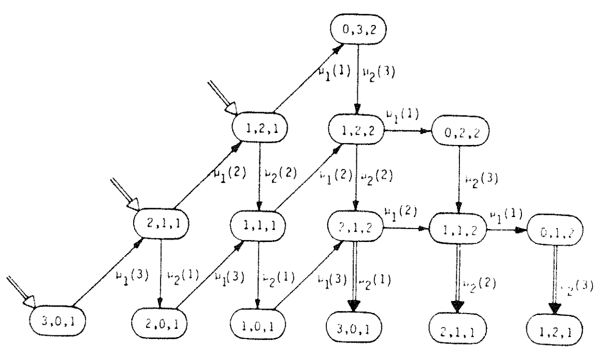

Fig. 4. State transition diagram for the cycle time for a cyclic queuing system with two stations and state-dependent service rates.

The arrows at the left hand side indicate the initial states a test customer meets on the arrival at station 1. The bold arrows at the lower right indicate the termination of the cycle time, i.e. the transitions into the absorbing space H.

The exact analysis of the cycle time df proceeds as follows:

- Definition of the set of $M\cdot(M+1)$ differential equations, one for each state of S according to Eq. (3.1).
- Calculation of the arrival state distribution for the arrival of a test customer at station 1 according to [17,18].
- Numerical solution of the system of differential equations either in the time domain (e.g. by a Runge-Kutta method) or in the Laplace-domain (i.e., solving for the corresponding eigenvalues).
- Calculation of the complementary df of the cycle time according to Eq. (3.3).

For larger populations M, larger queuing network structures, or both, the exact solution of the df requires an enormous amount of computing. Instead of the full df, we may be interested in the moments of lower order especially the second moment m_2 or the coefficient of variation c (the first moment is already known from the theory of product-form queuing networks [1,2]).

The solution for the ordinary moments requires only the solution of a set of linear equations according to Eq. (3.2), one set for each order k, k = 1,2,.... In the particular case of Fig. 4, each set can be solved recursively: starting with the right hand side column, all quantities are obtained by proceeding bottom-up, column-by-column from right to left.

From the lower conditional moments (or unconditional total moments), say k = 1,2 or k = 1,2, and 3, we can approximately construct the df. In Section 2 we will mainly concentrate on the coefficient of variation of cycle times only.

3.3 Approximate Analysis

To further reduce the computational amount, several approximate methods can be applied. Two different methods have been developed so far:

(a) Network transformations

(b) Independent flow time approximations

a) Network Transformations

Closed product-form queuing networks can be transformed into a cyclic queuing network of two stations (as shown in Fig. 3b) consisting of a considered station and a "composite station" by use of the so-called Norton Theorem for product-form queuing networks [3,5]. This transformation however, is only exact with respect to the throughput rate and state probabilities and, therefore, for the average cycle time. The higher moments and the df of the cycle time are only approximations.

This method has proved to be quite accurate for the following cases, see [15].

- Aggregation of several tandem stations into one composite station
- Aggregation of several parallel stations with equal flow times into one composite station.

By this method, certain types of networks can be structurally reduced step by step to facilitate the numerical computations. However, subnets with cyclic paths should not be aggregated into a composite station; although the average time is exact, the second moment can be quite underestimated.

b) Independent Flow Time Approximation (IFTA)

This method is applied to the original network structure. The flow times a test customer may observe in each of the passed stations are analyzed under an independently calculated arrival state distribution. The total flow time can be found by a summation of all component flow times, e.g., by convolutions of their respective df's.

First results of this method are quite encouraging [19]. This method will be reported by a forthcoming paper.

4. RESULTS AND VALIDATION

4.1 Test-Bed Networks

In the following a number of examples will be reported to show how priorities influence the cycle time in queuing networks and how accurate the analysis method works. The examples refer to three test-bed networks:

Test-Bed Network 1: Cyclic queuing network with two stations and two classes of nonpreemptive priorities (structure as in Fig. 3a).

Test-Bed Network 2: as 1, but with preemptive priorities.

Test-Bed Network 3: Central server queuing network model (CSM) with three stations and two classes of preemptive priorities (see Fig. 1).

4.2 Results

a) Results for Test-Bed Network 1

This model has been analyzed for a fixed population $M_1 = M_2 = 2$. The service rates are:

$\mu_{11} = 0.2$ μ_{12} variable

$\mu_{21} = 0.1$ $\mu_{22} = 0.1$

The results are given for the average cycle times t_{C1} and t_{C2} and for the cycle time coefficients of variation c_1 and c_2 for high priority and low priority customers, respectively (The average cycle times are exact values). All results are compared to simulations (given in brackets).

μ_{12}	t_{C1}	t_{C2}
1.0	27.35 (27.41 ± 0.15)	81.16 (81.71 ± 0.88)
0.1	30.08 (30.08 ± 0.11)	75.96 (76.27 ± 0.63)
0.01	109.4 (109.4 ± 1.13)	223.0 (223.4 ± 1.96)
0.001	903.5 (902.9 ± 8.10)	2023. (2024. ± 18.5)
0.0001	8806. (8802. ± 90.0)	20020 (20039 ± 212.)

μ_{12}	c_1	c_2
1.0	.6275 (.5846 ± .0028)	.7378 (.6434 ± .0044)
0.1	.6072 (.5763 ± .0023)	.6628 (.6300 ± .0048)
0.01	.8906 (.9111 ± .0074)	.6673 (.6266 ± .0076)
0.001	1.100 (1.097 ± .0060)	.7033 (.6957 ± .0057)
0.0001	1.129 (1.127 ± .0059)	.7067 (.7038 ± .0046)

Table 1. Results for test-bed network 1

b) Results for Test-Bed Network 2

In case of preemptive priorities, class 1-customers are not affected by class 2-customers. Therefore, for class 1 the results of the single-class network applies which are known explicitly [11,13, 15].

Subsequently, we give the results for class 2-customers for various populations. The service rate parameters are as in a).

b1) $M_1 = 2$, $M_2 = 2$

μ_{12}	t_{C2}	c_2
1.0	173.3 (173.4 ± 2.27)	.8468 (.7721 ± .0149)
0.1	190.9 (188.8 ± 1.84)	.7585 (.7059 ± .0176)
0.01	435.8 (435.4 ± 11.0)	.6196 (.6113 ± .0170)
0.001	3516. (3599. ± 239.)	.7076 (.7368 ± .0607)

b2) $M_1 = 2$, $M_2 = 5$

μ_{12}	t_{C2}	c_2
1.0	355.9 (357.0 ± 4.13)	.6756 (.6005 ± .0118)
0.1	367.0 (367.1 ± 5.98)	.6212 (.5652 ± .0066)
0.01	899.9 (902.7 ± 17.3)	.4532 (.4452 ± .0107)
0.001	8750. (8423. ± 459.)	.4504 (.4314 ± .0243)

b3) $M_1 = 5$, $M_2 = 2$

μ_{12}	t_{C2}	c_2
1.0	1665. (1729. ± 113.)	.9584 (.8998 ± .0463)
0.1	1684. (1710. ± 105.)	.9436 (.8696 ± .0355)
0.01	1896. (1928. ± 83.5)	.8170 (.7885 ± .0404)
0.001	4743. (4869. ± 304.)	.6151 (.5871 ± .0677)

b4) $M_1 = 5$, $M_2 = 5$

μ_{12}	t_{C2}	c_2
1.0	3245. (3407. ± 294.)	.8079 (.6667 ± .0295)
0.1	3257. (3345. ± 219.)	.7991 (.6665 ± .0351)
0.01	3436. (3408. ± 171.)	.6995 (.6470 ± .0371)
0.001	10020 (9716. ± 480.)	.4428 (.4596 ± .0532)

Table 2. Results for test-bed network 2

c) Results for Test-Bed Network 3

As in case b), the class 1-customers are not affected by class 2-customers. The results for the average cycle time are exactly known [1,2]; results for the cycle time coefficient of variation have been validated in [15], so that we can concentrate on the low priority class only.

We note, that the cycle time analysis has been based on the additional approximation of network transformation: both I/O-stations of the CSM for class 2-customers have been aggregated to one composite server. We consider only those cycles starting at station 1.

The parameters of the CSM are as follows:

$M_1 = 2$, $M_2 = 2$ and 5

$\mu_{11} = \mu_{21} = \mu_{31} = \mu_{22} = \mu_{32} = 1$, μ_{12} variable

$q_{12,1} = q_{12,2} = 0.5$

c1) $M_1 = 2$, $M_2 = 2$

μ_{12}	t_{C2}	c_2
10.	4.549 (5.757 ± 0.032)	.669 (.7218 ± .0078)
1.0	9.480 (10.76 ± 0.132)	.717 (.7277 ± .0099)
0.1	74.35 (73.07 ± 2.025)	.722 (.7254 ± .0265)
0.01	737.0 (730.4 ± 61.67)	.707 (.7353 ± .1026)

c2) $M_1 = 2$, $M_2 = 5$

μ_{12}	t_{C2}	c_2
10.	6.680 (4.873 ± 0.090)	.502 (.8804 ± .0159)
1.0	18.70 (19.43 ± 0.153)	.569 (.5969 ± .0096)
0.1	185.1 (185.4 ± 6.620)	.460 (.4748 ± .0164)
0.01	1842. (1805. ± 175.0)	.447 (.4433 ± .0406)

Table 3. Results for test-bed network 3

4.3 Validation

The presented analysis method for transit delays in priority queuing networks rests on various approximation hypothesises:

(1) decomposition of the priority network into networks without priorities

(2) aggregation of subnets into composite stations.

Approximation (1) has been carried out with respect to state distributions; the application of that principle to transit delay df's is one source of error.

Approximation (2) has been applied to reduce the complexity in the analysis. This method does work under certain conditions as mentioned in Section 3.3. A further reduction of errors will be obtained through the advanced analysis procedure IFTA.

For all these sources of errors, most analysis results are within a 10 % range of the simulation results which is good enough in most cases of applications. Variations in the population size, service rates, priority schedule, or network structure indicate that the analysis method applies to a wider range of applications.

CONCLUSION

This paper presents - for the first time - an analysis method for transit delay distributions of Markovian priority queuing networks. This problem could be solved exactly in principle using the same method as for single-class networks, see [15]. To reduce the computational complexity, the problem has been simplified by decomposition of the multi-class priority network into equivalent single-class networks and aggregation of subnetworks into composite server stations. The results indicate that the method applies to a wider range of parameters as population size, service rates, priority schedule ,or network structure. The method can be applied to quite realistic cases where priorities are used to meet specific real-time percentile criterions as in case of dial-tone or acknowledgement delays in telecommunication systems, or in case of turn-around delays in interactive computer systems.

ACKNOWLEDGEMENT

The programming and computing efforts of Mr. Stefan Hohl are greatly appreciated.

REFERENCES

[1] Baskett, F., Chandy, K.M., Muntz, R.R., Palacios, F.G., Open,Closed and Mixed Networks of Queues with Different Classes of Customers. J.ACM 22 (1975), 248-260.

[2] Reiser, M., Lavenberg, S.S., Mean Value Analysis of Closed Multichain Queuing Networks. J.ACM 27 (1980), 313-322.

[3] Chandy, K.M., Herzog, U., Woo, L., Parametric Analysis of Queuing Networks. IBM J. Res. and Develop. 19 (1975), 36-42.

[4] Kuehn, P.J., Approximate Analysis of General Queuing Networks by Decomposition. IEEE Trans. on Commun. COM-27 (1975), 113-126.

[5] Kritzinger, P.S., van Wyk, S. Krzesinski, A.E., A Generalization of Norton's Theorem for Multiclass Queuing Networks. Int. Journ. on Performance Evaluation 2 (1982), 99-107.

[6] Kaufman, J.S., Approximate Analysis of Priority Scheduling Disciplines in Queuing Network Models of Computer Systems. Proc. 6th Int. Conf. on Computer Commun., London (1982), 955-961.

[7] Morris, R.J.T., Priority Queuing Networks. Bell System Technical Journal 60 (1981), 1745-1763.

[8] Schmitt, W., Approximate Analysis of Markovian Queueing Networks with Priorities. 10th International Teletraffic Congress, Montreal (1983), Congressbook, paper 1.3-3.

[9] Schmitt, W., Traffic Analysis of Priority Queueing Networks. 39th Report on Studies in Congestion Theory, Institute of Switching and Data Technics, University of Stuttgart (1984).

[10] Schmitt, W., On Decompositions of Markovian Priority Queues and Their Application to the Analysis of Closed Priority Queuing Networks. Proc. Performance '84, Elsevier Science Publishers B.V., North Holland (1984), 393-407.

[11] Chow, We-Min, The Cycle Time Distribution of Exponential Cyclic Queues. J.ACM 27 (1980), 281-286.

[12] Daduna, H., Passage Times for Overtake-Free Paths in Gordon-Newell Networks. Adv. Appl. Prob. 14 (1982), 672-686.

[13] Schassberger, R., Daduna, H., The Time for a Round-Trip in a Cycle of Exponential Queues. J.ACM 30 (1983), 146-150.

[14] Boxma, O.J., The Cyclic Queue with one General and one Exponential Server, Adv. Appl. Prob. 15 (1983).

[15] Kuehn, P.J., Analysis of Busy Periods and Response Times in Queuing Networks by the Method of First Passage Times. Proc. Performance '83, North Holland Publ. Comp. (1983), 437-455.

[16] Syski, R., Markovian Queues. Symp. on Congestion Theory. The Univ. of North Carolina Press, Chapel Hill (1964), 170-227.

[17] Lavenberg, S.S., Reiser, M., Stationary State Probabilities at Arrival Instants for Closed Queuing Networks with Multiple Types of Customers. J. Appl. Prob. 17 (1980), 1048-1061.

[18] Sevcik, K.C., Mitrani, I., The Distribution of Queuing Network States at Input and Output Instants, J.ACM 28 (1981), 358-371.

[19] Hohl, S., Kuehn, P.J., Approximate Analysis of Flow Time Processes in Queueing Networks by the Method of Independent Flow Time Approximation. Monograph Univ. of Stuttgart (1985).