TD(05) 034

Meeting details

(4th MCM COST 290 Action, Wuerzburg, Germany, Oct 13-14 2005)

Evaluation of a Multi-server Delay System with a Generalized Poisson Input Stream

Seferin Mirtchev
Technical University of Sofia
8 Kliment Ohridski St.
1000 Sofia, Bulgaria
stm@tu-sofia.bg

Abstract

This paper deals with M(g)/M/n/k/S queue in witch we have generalized Poisson arrival process, exponential service time, multiple servers, limited waiting positions and finite number of customers. We use the generalized input Poisson stream that can be peaked, regular or smooth. The idea is based on the analytic continuation of the Poisson distribution and the classic Erlang's delay system M/M/n. We apply techniques based on birth and death process and state-dependent arrival rates. The influence of the peaked factor on the congestion probability, the mean system time and the waiting time distribution are studied. It is shown that the variance of the input stream changes significantly the characteristics of the delay systems. The advantages of simplicity and uniformity in representing both peaked and smooth traffic make this queue attractive in network analysis and synthesis.

Keywords

Queueing system, Poisson process, Peaked and smooth traffic

Working Group N 1

1. INTRODUCTION

Simple models like the classical single-server queues can often be used to obtain comprehensive results, e.g., to predict the global traffic behaviour. When modeling network traffic, packet and connection arrivals are often assumed to be Poisson processes because such processes have attractive theoretical properties.

Many studies on traffic measurements from a variety of communication networks, like Ethernet local area networks (LANs) and wide area networks (WANs) with Internet and asynchronous transfer mode (ATM), etc., have shown considerable difference between actual network traffic and assumptions in traditional theoretical traffic models.

Problems with the Poisson modeling are predicted in [11]. The authors are indicating that some arrivals deviate considerably from the Poisson distribution but user-initiated TCP session arrivals, such as remote-login and file-transfer, are well-modeled as Poisson processes with fixed hourly rates. The Internet traffic characteristics are studied by Cao et al [2]. They have shown that the arrivals tend to Poisson and the packet sizes tend to independence when the number of simultaneous transport connections increase.

Karagianis et al [7] believed that it is time to re-examine the Poisson traffic assumption in relation to the traffic carried within the Internet core. They have shown that the current network traffic can be well represented by the Poisson model for sub-second time scales.

The Poisson process is one of the simplest and most interesting stochastic processes [5,8]. One of its properties is that the mean and the variance of the number of calls within a time interval are equal. These features simplify analysis, but introduce inaccuracy.

The traffic flows inside a network are not Poissonion in general [1]. For many real teletraffic systems the mean number of events in an interval is not equal to the variance. The offered streams are said to be peaked or smooth according to whether the variance is bigger or smaller than the mean value, respectively.

The Bernouilli-Poisson-Pascal (BPP) method is used to approximate the main congestion functions associated with peaked and smooth traffic in lost-call-cleared systems. The BPP model represents peaked and smooth traffic by two separate models, and cannot represents arbitrary smooth traffic. The BPP method applies a state dependent arrival process with linear arrival rates [3].

The MMPP (Markov Modulated Poisson Process) traffic model that accurately approximates the long range dependence characteristics of Internet traffic traces is proposed in [10]. The Poisson Pareto burst process (PPBP) is presented in [12] as a simple and accurate model for aggregated Internet traffic.

Network analysis really requires a technique that can represent any kind of traffic, peaked or smooth, within the same model [3,6,9]. Most of the known methods are designed for a particular type of traffic, peaked or smooth, or, if they apply to both, do so using different models for different ranges of peakedness. The presented below method meets the above requirements.

In this paper peaked and smooth input streams are defined. They will be called a generalized Poisson process. A calculation method for the performance measures of a M(g)/M/n/k/S queue in witch we have generalized Poisson arrival process, exponential service time, multiple servers, limited waiting positions and finite number of customers is presented.

The idea is based on the analytic continuation of the Poisson distribution and the classic M/M/n system. We apply techniques based on birth and death process and state-dependent arrival rates.

2. GENERALIZED POISSON PROCESS

The Poisson process is a pure birth process with an arrival rate λ independent of the system state. The probability $P_i(t)$ of i arrivals in an interval whose duration is t seconds is given by

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t} . \tag{1}$$

Two more parameters, peakedness factor p and number of sources S is introduced for the generalized Poisson process. The process is said to be peaked, regular or smooth according to whether $p > 1$, $p = 1$ or $p < 1$, respectively.

Calls arrive in a generalized Poisson stream at rate λ_i which depends on the number of calls in the system. The time between successive call arrivals is exponentially distributed with different parameter λ_i. This generalized Poisson stream has memoryless property.

The arrival rate is state-dependent

$$\lambda_i = \lambda(i+1)^{1-1/p} . \tag{2}$$

The state probabilities $P_i(t)$ in the case of a generalized Poisson process are

$$P_i(t) = \frac{(\lambda t)^i / (i!)^{1/p}}{\sum_{j=0}^{S} (\lambda t)^j / (j!)^{1/p}} . \tag{3}$$

The mean value (the average number of arrivals in an interval of length t) is

$$M(t) = \sum_{i=1}^{S} i P_i . \tag{4}$$

The variance of the number of arrivals in an interval of length t is

$$V(t) = \sum_{i=0}^{S} [i - M(t)]^2 P_i(t) . \tag{5}$$

When $p = 1$ and $S \to \infty$, $M(t) = \lambda t$ and $V(t) = \lambda t$ i.e. it is a regular Poisson process.

3. GENERALIZED MULTI-SERVER DELAY SYSTEM - MODEL DESCRIPTION

Let us consider a multi server queue M(g)/M/n/k/S with a generalized Poisson input stream M(g), exponential service time M, number of servers n, limited waiting room k and number of sources S (S>k). This is a birth and death process and we can use the general solution, as given in [5], for the stationary probability of having j customers in the system

$$P_j = \frac{\prod_{i=0}^{j-1} \lambda_i / \mu_{i+1}}{1 + \sum_{v=1}^{k+1} \prod_{i=0}^{v-1} \lambda_i / \mu_{i+1}} \qquad j = 0,1,2,...,n+k \,. \tag{6}$$

This generalized delay system may be described by selecting the birth-death coefficient as follows

$$\begin{aligned} \lambda_j &= \lambda\ (j+1)^{1-1/p} && j = 0,1,2,...,n+k \\ \mu_j &= j\mu && j = 1,2,3,...,n \\ \mu_j &= n\mu && j = n, n+1,...,n+k \end{aligned} \,. \tag{7}$$

The arrival rate is state-dependent and depends on the peakedness factor *p*. This limited delay system is always ergodic. The finite state-transition diagram is shown in Fig.1.

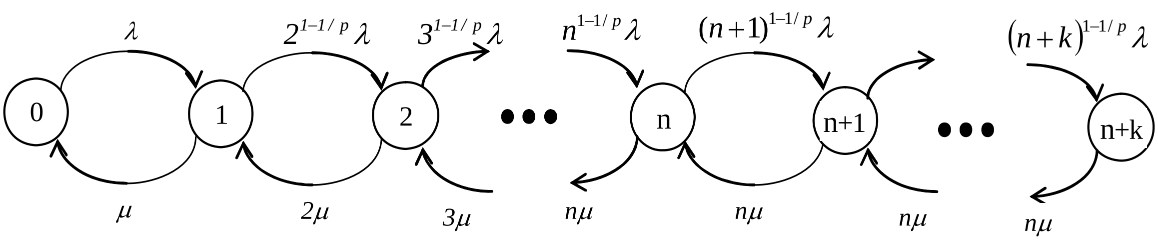


Fig.1. A state-transition diagram - M(g)/M/n/k/S queue

Applying these coefficients to the general solution of the birth and death process and using traffic intensity $a = \lambda/\mu$ we obtain the steady state probabilities

$$\begin{aligned} P_j &= \frac{a^j}{(j!)^{1/p}} P_0 && 0 \le j \le n \\ P_j &= \frac{a^j}{(j!)^{1/p}} \frac{j!}{n! n^{j-n}} P_0 && n \le j \le n+k \,. \\ P_0 &= \frac{1}{\sum_{i=0}^{n} \frac{a^i}{(i!)^{1/p}} + \sum_{i=n+1}^{n+k} \frac{a^i}{(i!)^{1/p}} \frac{i!}{n! n^{i-n}}} \end{aligned} \tag{8}$$

The offered traffic is calculated by means of the average arrival rate and the mean holding time

$$A = \bar{\lambda} \frac{1}{\mu} = \frac{1}{\mu} \sum_{j=0}^{n+k} \lambda_j P_j = a \sum_{j=0}^{n+k} (j+1)^{1-1/p} P_j \,. \tag{9}$$

The carried traffic is equivalent to the average number of busy servers

$$A_c = \sum_{i=0}^{n} i\,P_i + n \sum_{i=n+1}^{n+k} P_i \ . \tag{10}$$

4. GENERALIZED ERLANG DISTRIBUTION

Assume that the number of the servers is equal to the number of the sources. In this case the system has not any losses and delay, the whole offered traffic is carried and it is called the intended traffic load.

The stationary probability of having *j* customers in the system has generalized Erlang distribution

$$P_j{}' = \frac{a^j \big/ (j!)^{1/p}}{\sum_{i=0}^{S} a^i \big/ (i!)^{1/p}} \qquad j = 0,1,2,\dots,S \ . \tag{11}$$

The intended traffic is the equilibrium number of busy servers

$$A_i = \sum_{j=1}^{S} jP_j{}' \ . \tag{12}$$

The variance of the intended traffic is

$$V(A_i) = \sum_{j=0}^{S} (j - A_i)^2 P_j{}' \ . \tag{13}$$

The peakedness of the intended traffic is the variance to mean ratio

$$z = \frac{V(A_i)}{A_i} . \tag{14}$$

5. Mg/M/n/k/S SYSTEM – TRAFFIC CHARACTERISTICS

BLOCKING PROBABILITY. The time congestion probability B_t describes the fraction of time that all waiting rooms are busy

$$B_t = P_{n+k} \ . \tag{15}$$

The call congestion probability B_c is ratio of lost traffic (offered minus carried traffic) to offered traffic

$$B_c = \frac{A - A_o}{A} . \qquad (16)$$

WAITING PROBABILITY. The waiting probability is denoted by *P(>0)* which means that the waiting time probability is greater than *0*

$$P(>0) = 1 - \sum_{i=0}^{n-1} P_i - P_{n+k} . \qquad (17)$$

MEAN NUMBER OF CALLS. The mean number of calls present in the system in steady state by definition is

$$L = \sum_{j=1}^{n+k} jP_j . \qquad (18)$$

MEAN SYSTEM TIME. From the Little formula, we have the mean system time

$$T = \frac{L}{\bar{\lambda}} = L \Big/ \sum_{j=0}^{n+k} \lambda_j P_j . \qquad (19)$$

WAITING TIME DISTRIBUTION. Let us assume the first-come-first-out (FIFO) discipline. The waiting time distribution function *P(>t')* is defined as the probability of waiting time exceeding *t'*. From the probability theory it is given by

$$P(>t') = \sum_{i=n}^{n+k-1} P_i Q_i(>t') . \qquad (20)$$

An arbitrary call enters service when *i* calls are in the system P_i. Since the service time is exponentially distributed the probability that *i* calls terminate in time *(0,t']* becomes a Poisson distribution with mean $\mu t'$

$$Q_i(t') = \frac{(n\mu t')^i}{i!} e^{-n\mu t'} . \qquad (21)$$

The conditional probability $Q_i(>t')$ that the arbitrary call has to wait longer than *t'*, given *i* calls in the system, is expressed by:

$$Q_i(>t') = \sum_{r=0}^{i-n} \frac{(n\mu t')^r}{r!} e^{-n\mu t'} . \qquad (22)$$

6. CALCULATION OF THE STATE PROBABILITY

The traffic intensity *a* is not equal to the intended traffic in a case of a generalized Erlang process because we calculate the power of the Erlang unsymmetrical distribution. That is why we have to calculate the intended traffic A_i and the peakedness z when defining the traffic intensity a and peakedness factor p.

From the practical point of view we first define the intended traffic A_i and the peakedness z and after that calculate the traffic intensity a and peakedness factor p.

A fundamental question about the system defined by Eqs. (11), (12) and (14) is whether there exist solutions a, p for an arbitrary A_i, z. Although no formal proof seems to exist, this seems to be the case and the solution appears to be unique.

We can find solutions of the above system with the iterating method of consecutive replacements.

7. NUMERICAL RESULTS

In this section we give numerical results obtained by a Pascal program on a personal computer. The described methods were tested on a computer over a wide range of arguments.

Figure 2 shows the generalized Erlang distribution where the intended traffic is A_i = *10 erl*, the number of the sources S = *100* and the peakedness z is change from 0.6 to 1.5. It will be seen that when the peakedness z increases the probability distribution becomes broad about the mean.

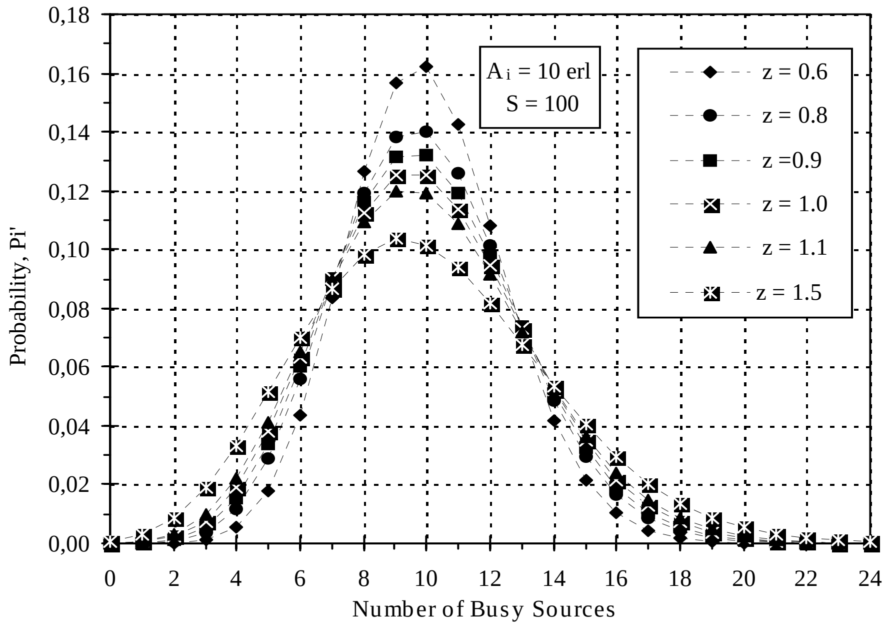


Fig.2. Generalized Erlang distribution when the intended traffic A_i = *10 erl*, the number of the sources S = *100* and different peakedness

Figures 3 and 4(a, b) illustrate the stationary probability distribution in a multi-server queue M(g)/M/n/k/S with a generalized Poisson input stream, 10 servers, 30 waiting rooms, 100 sources and different intended traffic A_i and peakedness z. It will be seen that when the utilization is from 0.9 to 1 erl and the peakedness is bigger than one the probabilities increase when the number of the customers in the system increases.

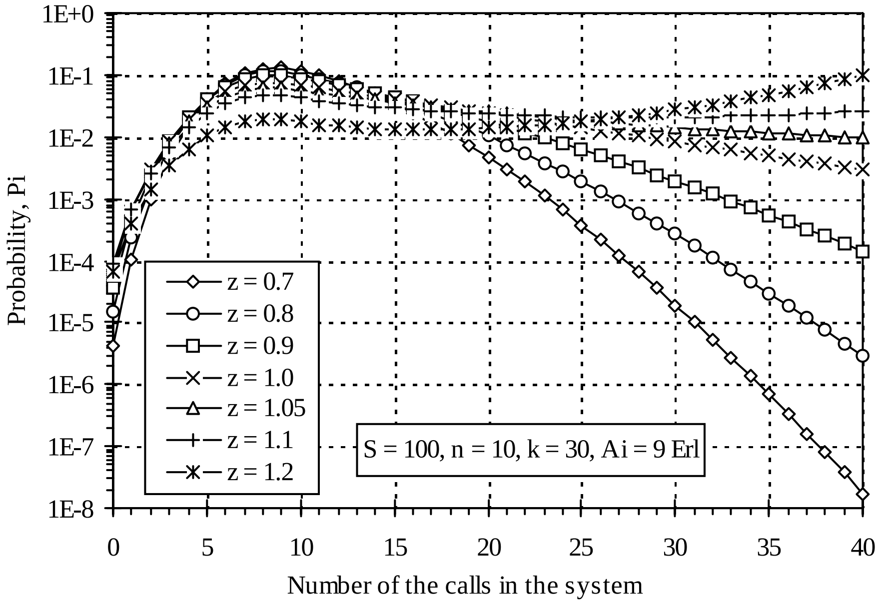
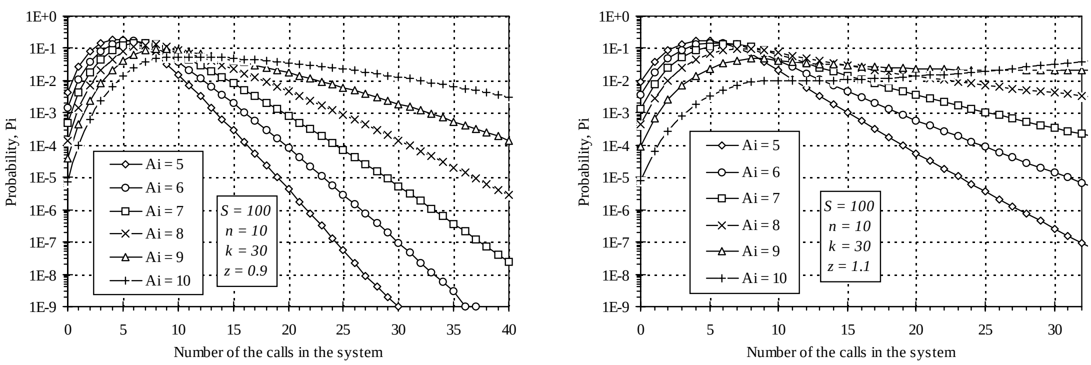


Fig.3. Stationary probability distribution in a multi-server queue with different peakedness of the intended traffic



a) b)

Fig.4. Stationary probability distribution in a multi-server queue with a smooth – a) z = 0.9 and peaked input stream – b) z = 1.1

Figures 5 and 6 show the call and time congestion probabilities in a multiple delay system with 10 servers, 100 sources, 0.9 erl intended traffic and different peakedness as function of the buffer size. When the utilization is high (0.9 - 1 erl) and the input stream is peaked (z = 1.05 – 1.2) the influence of the buffer size of the congestion probability is negligible. We have to notice that the offered traffic is bigger than intended in the M(g)/M/n/k/S queue when the input stream is peaked.

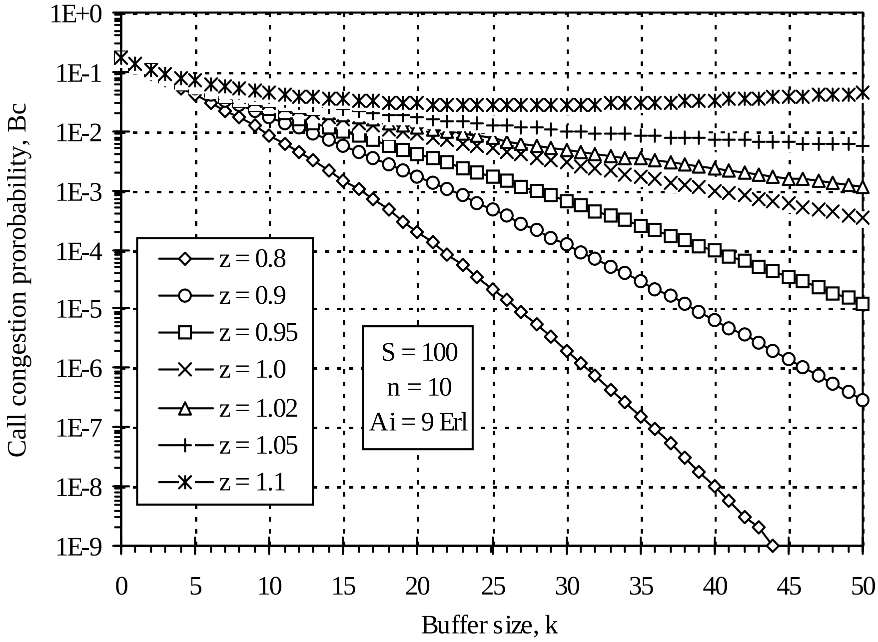


Fig.5. Call congestion probability in a multiple delay system with different peakedness of the intended traffic

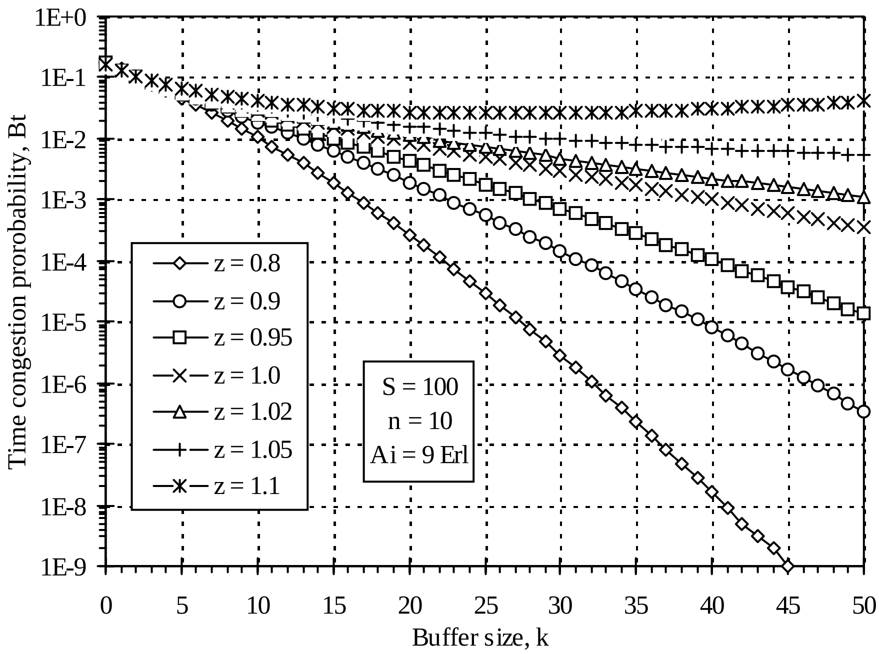
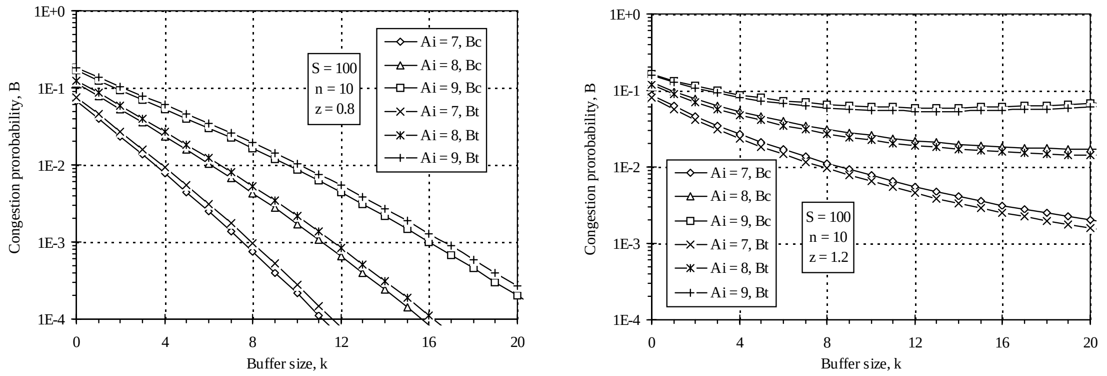


Fig.6. Time congestion probability in a multiple delay system with different peakedness of the intended traffic

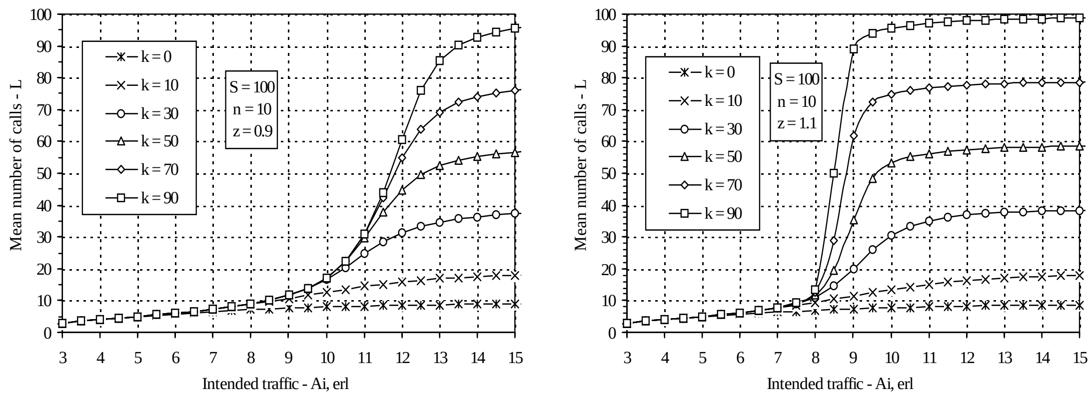
Figure 7 (a, b) compares the call and time congestion probabilities in a multiple delay system with 10 servers, 100 sources and different intended traffic and peakedness as function of the buffer size.

Figure 8 (a, b) presents the mean number of calls in the system *L* and figure 9 (a, b) presents the normalized mean system time ($W' = W/\tau$) as function of the intended traffic when the number of servers is 10, the number of sources is 100, the peakedness is 0.9 and 1.1 respectively and different waiting room.



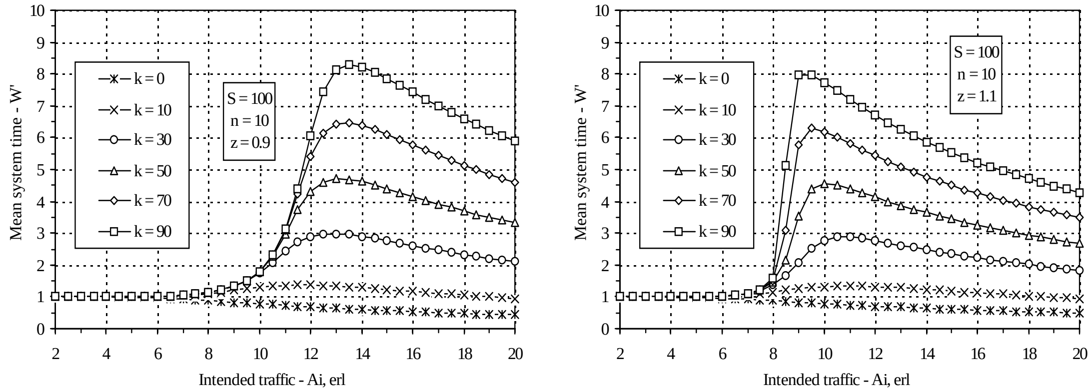
a) b)

Fig.7. Call and time congestion probabilities in a multiple delay system with different intended traffic when the input stream is smooth – a) z = 0.8 and peaked – b) z = 1.2



a) b)

Fig.8. Mean number of the calls in the system when the input stream is smooth – a) z = 0.9 and peaked – b) z = 1.1

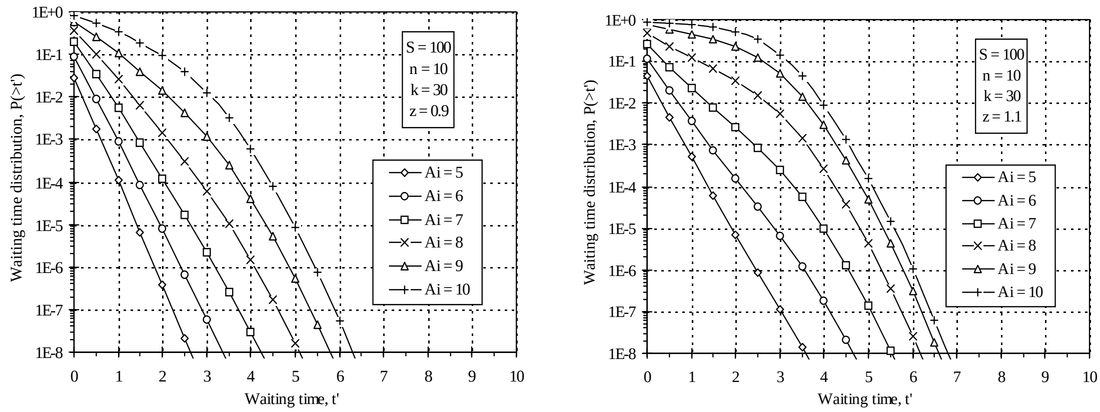


a) b)

Fig.9. Normalized mean system time when the input stream is smooth – a) z = 0.9 and peaked – b) z = 1.1

Figure 10 (a, b) illustrates the waiting time distribution as function of the normalized waiting time when the number of servers is 10, the number of sources is 100, the buffer size is 30, the peakedness is 0.9 and 1.1 respectively and different intended traffic.

It is shown that the influence of the peakedness over the performance measures is significantly.



a) b)

Fig.10. Waiting time distribution as a function of the normalized waiting time when the input stream is smooth – a) z = 0.9 and peaked – b) z = 1.1

8. CONCLUSIONS

In this paper a generalized Poisson process is introduced and evaluated. A basic model for a queueing system Mg/M/n/k/S is examined in detail.

The proposed method provides a unified framework to model peaked and smooth traffic. Numerical results and subsequent experience have shown that this method is accurate and useful in both analyses and simulations of teletraffic systems.

The importance of a multiple delay system in a case of a generalized Poisson input stream comes from its ability to describe behaviour that is to be found in more complex real queueing systems. It is the case in a general traffic system, which is an important feature in designing telecommunication systems.

In conclusion, we believe that the presented generalized Poisson process and queueing system will be useful in practice. As part of future work, we plan to analyzed a processor sharing system with a generalized Poisson input stream.

REFERENCES

1. Akimaru H. and K. Kawashima. Teletraffic Theory and Applications, Springer-Verlag, 1993.
2. Cao J.,W. Cleveland, D. Lin, and D. Sun. Internet Traffic Tends to Poisson and Independent as the Load Increases. Bell Labs Technical Report, 2001.
3. Delbrouck L.E.N. A Unified Approximate Evaluation of Congestion Functions for Smooth and Peaky Traffics, IEEE Trans. on Commun., Vol.29, No 2, 1981, pp.85-91.
4. Girard A. Routing and Dimensioning in Circuit-Switched Networks, Addison-Wesley, 1990.
5. Iversen, V.B. Teletraffic Engineering Handbook, ITU-D & ITC. 312 pp. Edition spring 2004, http://www.com.dtu.dk/education/34340/.
6. Iversen V. and S. Mirtchev. Generalized Erlang Loss Formula, Electronics Letters, Vol. 32, No: 8, April 1996, pp.712-713.
7. Karagiannis T., M. Molle, M. Faloutsos, A. Broido, A Nonstationary Poisson View of Internet Traffic, IEEE INFOCOM 2004, Vol.23,no.1, March 2004, pp.1559-1570.
8. Kleinrock L. Queueing Systems, Volume I: Theory, John Wiley & Sons, 1975.
9. Mirtchev S. and I. Stanev. Evaluation of a Single Server Delay System with a Generalized Poisson Input Stream, ITC19, Beijing, China, Vol.6a, 2005, pp.553-542.
10. Muscariello L., M.Mellia, M.Meo, M.Ajmone Marsan, R. Lo Cigno, An MMPP-Based Hierarchical Model of Internet Traffic, IEEE ICC 2004, Vol.27, no.1, June 2004, pp.2143-2147.
11. Paxson V. and S. Floyd. Wide Area Traffic: The Failure of Poisson Modelling, IEEE/ACM Transactions on Networking, Vol. 3, no.3, June 1995, pp. 226-244.
12. Zukerman M., T. Neame and R. Addie, Internet Traffic Modelling and Future Technology Implications, IEEE INFOCOM 2003, Vol.22, no.1,March 2003, pp.578-596.