A Discrete Time Queueing Model with a Constant Packet Size

Seferin T. Mirtchev[1] and Rossitza Goleva[2]

Abstract – **The importance of IP network will further increase and it will serve as a platform for more and more services, requiring different types and degrees of service quality. Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. In this paper, we investigate the queueing behaviour found in IP output buffers. This queueing increases because multiple streams of packets with different length are being multiplexed together. To analyze these types of behaviour, we study the discrete-time version of the "classical" queue model M/D/1/k called Geo/D/1/k. This is a discrete-time single server FIFO queue with Bernoulli arrivals and deterministic service times. We develop balance equations for the state of the system, from which we derive packet loss and delay results.**

Keywords – **Queueing system, Queueing analyses, Discrete time queue, IP traffic modelling**

I. Introduction

The initial motivation for this paper is the necessity of traffic engineering in NGN networks. Many analyses of Internet traffic behaviour require accurate knowledge of the traffic characteristics for purposes ranging from a management of the network quality of service to modelling the effect of new protocols on the existing traffic mix.

Modern architectures and protocols are being standardized, which aims at guaranteeing the quality of service delivered to users. The proper functioning of these protocols requires an increasingly detailed knowledge for statistical characteristics of IP packets. The amount of information flowing through the network also increases, and the challenge is to obtain the accurate information from a huge set of data packets [5,7]. The packet queueing in an IP router arises because multiple streams of packets from different input ports are being multiplexed together over the same output port [9,14].

Many communication systems operate on a discrete-time basis and events can only happen at regularly spaced epochs.

Discrete-time queueing systems have been receiving increased attention in recent years due to their usefulness in modeling and analyzing various types of communication systems. But the technical difficulties which arise in working with a discrete time scale are considerably greater than in dealing with continuous time models although the mathematics behind is more elementary than in the continuous time case. The reason for this is the combinational complexity which appears in the solution procedures for the global balance equations of these systems.

Discrete-time queueing systems have been a research topic for several decades now and there are many reference works on discrete-time queueing theory. Over the years, different methodologies have been developed to assess the performance of queueing systems. The two main analytical approaches are the matrix analytic method and the transform method for discrete and for continuous-time analyses. Many authors have considered the Geo/G/1 discrete-time queueing system [8], [10], [12], [13].

[6] has studied a discrete-time Geo/D/1 and Geo/D/1/n queues. The closed-form expressions for the steady-state distributions of the queue length and of the unfinished work in system (i.e. waiting time) are obtained by the method of analysis, using Lindley's equation.

In [1] a complete study of a discrete-time single-server queue with geometrical arrivals of both positive and negative customers is carried out. Negative arrivals are used as a control mechanism in many telecommunication and computer networks. [2] has concerned with the study of a discrete-time single-server retrial queue with geometrical inter-arrival times and a phase-type service process. An iterative algorithm to calculate the stationary distribution of Markov chain is given.

[11] has proposed a traffic model and a parameter fitting procedure that are capable of achieving accurate prediction of the queuing behaviour for IP traffic exhibiting long-range dependence. The modelling process is a discrete-time batch Markovian arrival process (dBMAP) that jointly characterizes the packet arrival process and the packet size distribution. In the proposed dBMAP, packet arrivals occur according to a discrete-time Markov modulated Poisson process (dMMPP) and each arrival is characterized by a packet size with a general distribution that may depend on the phase of the dMMPP.

[3] has presented an introduction to bandwidth estimation and a solution to the problem of the best-effort traffic for the case when the quality criteria specify negligible packet loss. The solution is a simple statistical model, which is built and validated using queueing theory and extensive empirical study.

It has been shown [4] that in the case of real-time communications, for which small buffers are used for delay reasons, short range dependence dominates the loss process and so the Markov-modulated Poisson process (MMPP) might be a reasonable source model. They have presented an exact mathematical model for the loss process of a MMPP+M/Ek/1/K queue and have concluded that the packet size distribution affects the packet loss process and thus the efficiency of forward error correction.

[1]Seferin Mirtchev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: stm@tu-sofia.bg .

[2]Rossitza Goleva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, e-mail: rig@tu-sofia.bg .

In this paper, we investigate the basic queueing behaviour of packets found in IP output buffers. The traffic is being generated from the packets of constant size that arrive for transmission on the link. The packets can queue up and loss if their size is bigger than the free positions of the buffer. The quality metrics for the best-effort traffic on the Internet are the packets loss and delay. To analyze these types of behaviour, we study the discrete-time version of the “classical” queue model M/D/1/k called Geo/D/1/k. We developed balance equations for the state of the system, from which we derived packets loss and delay.

II. Balance Equations - Geo/D/1/k Queue

We investigate a single server finite queue delay system Geo/D/1/k with a geometric distributed inter-arrival time and a constant packet length. We consider queueing phenomena in discrete-time queueing systems. We assume a fundamental time unit (time slot), the time to transmit an octet (byte), T_b. Customers arrive in the queueing system under consideration during the consecutive slots, but they can only start service at the beginning of slots. That is, service of customers is synchronized with respect to slot boundaries. Further, customer service times are integer multiples of the slot length, which implies that customers leave the system at slot boundaries. During the consecutive slots, packets arrive in the system, are stored in a finite capacity queue and are served by a single server on a first in first out (FIFO) basis (Fig. 1).

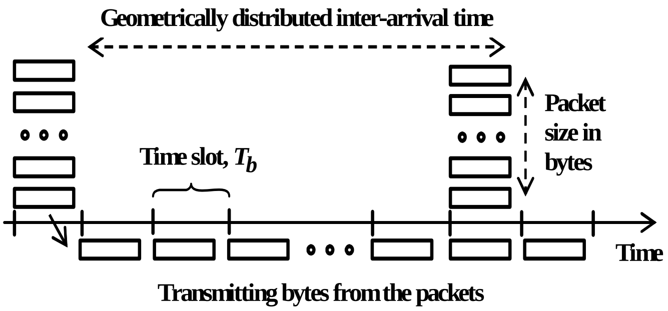

Fig. 1. Timing of events in the Geo/D/1/k queue.

We use a Bernoulli process for the packet arrivals, i.e. a geometrically distributed number of slots between arrivals. Let the probability that a packet arrives in an octet slot is $\boldsymbol{p}$.

Thus we have a batch arrival process with geometrically distributed inter-arrival times. That is, the number of slots that separate consecutive slots where there are customer arrivals, constitute a series of independent and identically geometrically distributed random variables. The probability of no octets arriving in a time slot is

$$a_0 = 1 - p \tag{1}$$

In this model, we assume a constant packet size with a value $\boldsymbol{m}$.

The probability that $\boldsymbol{m}$ octets arrive in a time slot is

$$a_m = p \tag{2}$$

The mean packet service time is the octet transmission time multiplied by the number of octets

$$\tau = T_b\, m, \ \ s \tag{3}$$

The mean arrival rate is

$$\lambda = p/T_b\,, \quad packets/s \tag{4}$$

Therefore, the offered traffic is given by

$$A = \lambda\tau = p\ m, \ \ Erl \tag{5}$$

The average inter-arrival time of the packets in time slots of the geometric distribution is

$$m_o = 1/p \tag{6}$$

The variance of the inter-arrival time of the packets is

$$v_i = (1-p)/p^2 \tag{7}$$

We define the state probability $\boldsymbol{P_i}$ of being of state $\boldsymbol{i}$, as the probability that there are $\boldsymbol{i}$ octets in the system at the end of any time slot. For the system to contain $\boldsymbol{i}$ bytes at the end of any time slots it could contain any of *0, 1, 2, ..., i+1* at the end of the previous slot. State $\boldsymbol{i}$ can be reached from any of the states $\boldsymbol{0}$ up to $\boldsymbol{i}$ by a precise packet arrival. To move from $\boldsymbol{i+1}$ to $\boldsymbol{i}$ there should be no arrivals.

We can write the first equation by considering all the ways in which it is possible to reach the empty state

$$P_0 = (P_0 + P_1)a_0 \tag{8}$$

Similarly, we find a formula for the next state probabilities by writing the balance equations

$$P_i = P_{i+1}a_0\,, \qquad 1 \le i \le m-1 \tag{9}$$

We continue with this process and take into account that it is possible to enter a packet in a time slot with length $\boldsymbol{m}$ bytes

$$\begin{aligned}
P_m &= (P_0 + P_1)a_m + P_{m+1}a_0 \\
P_{m+1} &= P_2 a_m + P_{m+2}a_0 \\
P_{m+2} &= P_3 a_m + P_{m+3}a_0 \\
&\quad o \ o \ o \\
P_{k-m+1} &= P_{k-2m+2}a_m + P_{k-m+2}a_0 \\
P_{k-m+2} &= P_{k-2m+3}a_m + P_{k-m+3} \\
&\quad o \ o \ o \\
P_{k-1} &= P_{k-m}a_m + P_k \\
P_k &= P_{k-m+1}a_m + P_{k+1}
\end{aligned} \tag{10}$$

Then using the fact that all the state probabilities must sum to 1 we write the last equation

$$\sum_{i=0}^{k+1} P_i = 1 \tag{11}$$

We can solve the system Eqs. (8), (9), (10) and (11) and calculate the state probabilities.

III. Performance Measures

The carried traffic is equivalent to the probability that the system is busy

$$A_o = 1 - P_0\,,\ erl \tag{12}$$

The packet congestion probability is the ratio of lost traffic (offered minus carried traffic) to offered traffic

$$B = (A - A_o)/A \tag{13}$$

The mean number of bytes and packets present in the system in steady state by definition is

$$L_b = \sum_{j=1}^{k+1} j P_j\,,\ bytes; \qquad L_p = L_b/m\,,\ packets \tag{14}$$

From the Little formula, we have the normalized mean system time of the bytes (time is measured in time slots)

$$\frac{W_b}{T_b} = \frac{L_b}{T_b \lambda m} = \frac{L_b}{A} \tag{15}$$

IV. 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.

Fig. 2 shows the stationary probability distribution in a single server queue *Geo/D/1/k* with 0.95 erl offered traffic, 100 waiting positions and different packet length. We can see that the probability distributions are almost linear decreasing in logarithmic scale and the influence of the packet length distribution kind on the stationary probability is significant.

Fig. 3 compares the packet congestion probability as function of the traffic intensity when the waiting positions are 200 and different packet size. We can see that the influence of the packet length on the packet congestion probability is big.

Fig. 4 presents the normalized mean system time of the bytes (W/T_b) as function of the traffic intensity when the queue length is 200 bytes and different packet size. The influence of the packet size on the mean system time is significant when the offered traffic is smaller then 1 erl.

We compare the discrete-time queue *Geo/D/1/k* with continuous time Poisson arrival queue model *M/D/1/k*. In this model we can accept that the packet length is one byte. When the packets size increases (3, 5 and 10 bytes) the stationary probability distribution is change and the packet congestion probability and mean system time increase vastly.

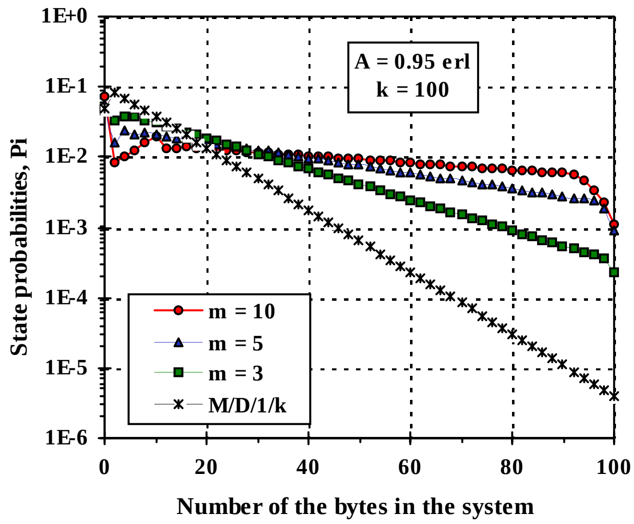

Fig. 2. Stationary probability distribution

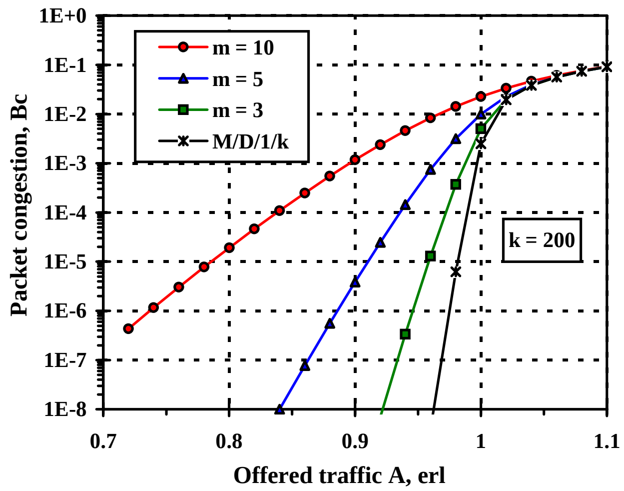

Fig. 3. Stationary probability distribution

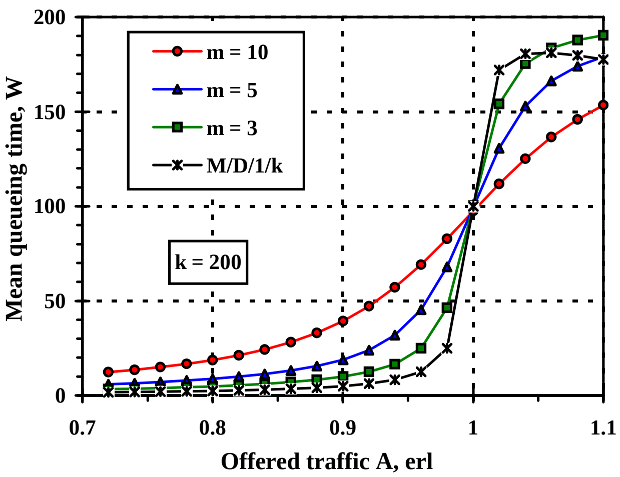

Fig. 4. Stationary probability distribution

V. Conclusion

In this paper a basic discrete-time single server teletraffic system Geo/D/1/k with a constant packet length is examined in detail. We present an analysis of this discrete-time queue.

The proposed approach provides a unified framework to model discrete-time single server queue. Numerical results and subsequent experience have shown that this approach is accurate and useful in both analyses and simulations of traffic systems.

The importance of a single server queue in a case of a geometric input stream and constant packet length 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 networks and systems.

The results presented here add a new aspect to the evaluation of the discrete-time queueing system, and serve as a basis for future research on guaranteeing the quality of service

In conclusion, we believe that the presented analytical model will be useful in practice.

Acknowledgement

This paper is sponsored by the Ministry of Education and Science of the Republic of Bulgaria in the framework of project ВУ-ТН-105/2005 “Multimedia Telecommunications Networks Planning with Quality of Service and Traffic Management”.

References

[1] Atencia I. and P. Moreno. A single-server G-queue in discrete-time with geometrical arrival and service process. Perform. Eval. 59: pp. 85-97 (2005)

[2] Atencia I, P. Bocharov and P. Moreno. A discrete-time Geo/PH/1 queueing system with repeated attempts. Информационные процессы, Том 6, N: 3, стр. 272-280 (2006).

[3] Cao J., W. Cleveland and D. Sun. Bandwidth estimation for best-effort Internet traffic. Statist. Sci., Volume 19, Number 3 (2004), pp. 518-543.

[4] Dan G., V. Fodor, and G. Karlsson. “Packet size distribution: an aside?” in Proc. of QoS-IP’05, pp. 75–87, February 2005.

[5] Farber J., S. Bodamer and J. Charzinski. Measurement and Modelling of Internet Traffic at Access Networks. Proceedings of the EUNICE'98,1998,196-203.

[6] Gravey A., J. Louvion and P. Boyer. On the Geo/D/1 and Geo/D/1/n queues. Performance Evaluation, 11, pp. 117 —. 125, 1990.

[7] Janevski T., D. Temkov, A. Tudjarov. Statistical Analysis and Modelling of the Internet Traffic. ICEST Sofia, 2003, pp. 170-173.

[8] Mirtchev S., G. Balabanov and S. Statev. New Teletraffic Models in the IP Networks. National Conference with Foreign Participation, Telecom'2006, Varna, Bulgaria, 2006 (in Bulgarian).

[9] 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.

[10] Pitts J.and J. Schormans. Introduction to IP and ATM Design and Performance - 2nd Ed. John Wiley & Sons, 2000.

[11] Salvador P., A. Pacheco and R. Valadas. Modelling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs. Computer Networks, Volume 44, Issue 3 , 2004, pp. 335-352.

[12] Vicari N. and P. Tran-Gia. A numerical analysis of the Geo/D/N queueing system. Technical Report 04, COST-257, 1996.

[13] Zhang Z. and N.Tian. Discrete Time Geo/G/1 Queue with Multiple Adaptive Vacations. Queueing Systems, Volume 38, Number 4, August 2001, pp. 419-429.

[14] 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.