TOP (2014) 22:454–457
DOI 10.1007/s11750-014-0328-x

DISCUSSION

Comments on: Queueing models for the analysis of communication systems

Alexander Dudin

Published online: 10 May 2014

In the article, authors restrict their attention to the class of discrete-time single-server queueing models useful for performance analysis of communication systems. This is quite natural because the leader of the author's group professor Herwig Bruneel has the great experience in analysis of discrete-time queues, see his book Bruneel and Kim (1993) published twenty years ago and an impressive amount of journal and conference papers published by the authors after this book. As the authors mention, queueing models in discrete-time are very appropriate to describe traffic and congestion phenomena in digital communication systems, since these models reflect in a natural way the synchronous nature of modern transmission systems, whereby time is segmented into intervals ("slots") of fixed length and information packets are transmitted at slot boundaries only, i.e., at a discrete sequence of time points.

Because it is well recognized that correlation is a typical feature of the traffic in modern telecommunication networks; authors concentrate their efforts in this paper on analysis of queues with correlation in the arrival process. Taking into account popularity of so-called BMAP (Batch Markov Arrival Process), see Chakravarthy (2001); Lucantoni (1991), for modeling correlated traffic in continuous-time queueing systems, see, e.g., Heyman and Lucantoni (2003); Klemm et al. (2003), the authors might use the discrete counterpart of the BMAP, so-called DBMAP (Discrete Batch Markov Arrival Process) for description of the arrival process in their models. However, as authors mention in the text, in the case of the DBMAP usually it is possible to develop

This comment refers to the invited paper available at doi:10.1007/s11750-014-0330-3.

A. Dudin (✉)
Faculty of Applied Mathematics and Computer Science, Belarusian State University,
4, Nezavisimosti Ave., Minsk 220030, Belarus
e-mail: dudin@bsu.by
e-mail: dudin_alex@yahoo.com

efficient numerical procedures to compute the important performance characteristics of the queueing system, but not to get attractive analytical results. Because the authors had the aim to investigate the queues with correlation in the arrival process analytically, they consider another way to account possible correlation in the arrival process. They consider a very interesting so-called a session-based arrival model.

A session is more or less new notion for arrivals description in queueing theory. The main three known notions are: (i) a *customer*, or packet, request, etc., as an entity which arrives to the queueing system for the service; (ii) a *flow* as an infinite sequence of customers arriving to the queueing system at random time moments; (iii) a *batch* as a group of customers simultaneously arriving within a flow. Usually, it is assumed that a batch consists of a random number of customers. A session also consists of a random number of customers. The difference between batch and session consists of the following. All customers belonging to the same batch arrive to the system at the same moment while the customers belonging to a session arrive to the system not simultaneously, but one-by-one after random time intervals. Alternatively, the notion of a session may be defined as a modification of the notion of a flow. A flow is as an infinite sequence of customers. It does not have initiation and termination points. A session is a finite sequence of customers. It starts and ends at some random moments.

Mathematical description of a session in discrete-time settings is more or less simple. Basic definition supposes that, once a session starts at a some slot, strictly positive random number of customers belonging to this session arrives at each consequent slot until the session ends. So, the session is completely defined by distributions of two strictly positive random variables: the number of slots between the start and the end of a session and the number of customers from this session arriving in an arbitrary slot. Sure, it is possible to think about the case when these two random variables are dependent, but this would essentially complicate the analysis. In some works of authors, it is allowed to have gaps between the slots, in which customers from a session arrive. Aiming to provide a nice strict analytical analysis, in this paper authors consider in depth a special case of a session arrival, so-called train arrival. The train arrivals assume that exactly one customer from an active session arrives at each sequential slot.

In this article, the authors give good overview of the existing relevant results and implement very impressive and comprehensive analysis of a single-server queue with infinite buffer space and train arrival of two types of customers. One type has a priority over another type. The number of customers of both types generated during a slot may be correlated. Using a quite involved technique of multi-dimensional probability generating functions, authors analyze the packet delays of both types of customers. Analysis is implemented in several steps (computation of the multi-dimensional probability generation functions of the system states at the beginning of a random slot and at an arbitrary arrival instant, of the virtual and actual delay, moments of delay and probability tail of delay for non-priority type-2 customers, and, then, derivation of results for priority type-1 customers based on results obtained for type-2 customers). Essential technical difficulties, e.g., necessity of solution of a functional equation for the multi-dimensional probability generating functions, necessity of using and analyzing the roots of equations, etc. arise. But they are successfully bridged over by the authors. In particular, it was helpful to supplement analytical derivations by the probabilistic arguments.

Results of this paper are sustainable and hardly can be essentially complemented in a brief note. The presented literature survey may be supplemented as follows. This paper is devoted to analysis of a discrete-time model with session arrivals. Although discrete-time models are extremely useful for analysis of processes of information transmission in communication networks on low layers of Open Systems Interconnection model, consideration of higher layers sometimes should be made by means of continuous-time queueing models. In particular, it may be useful to exploit continuous-time queueing models with session arrivals for analysis of the process of the information transmission at network, transport and session layers. These queueing models are not well known in the literature. So, here we present a very short overview of existing in literature continuous-time queueing models with session arrivals.

It is well known that continuous-time Markov chains are a bit more complicated subject of research than the discrete-time Markov chains due to the necessity of account of time spent by the chain at each state until the next transition. Correspondingly, the basic definition of a discrete-time model with session arrivals is more simple comparing to the definition of a continuous-time model with session arrivals. In continuous-time, due to the random length of inter-arrival times of customers within a session, it is necessary to permanently explicitly specify whether a tagged session is still in a progress or it is ended.

Good verbal description of a continuous-time session arrival mechanism was first given in the paper Kist et al. (2005) where this mechanism was analyzed by means of a computer simulation and a problem of restricting the number of sessions, which can be established simultaneously, was considered.

To the best of my knowledge, the first paper where a continuous-time session arrival mechanism was described in language of the queueing systems was Lee et al. (2007). The session is a set of customers, which will sequentially arrive to the system, if the session is admitted to the system. The number of customers in the session is not known at the session arrival moment. After arrival of the session (and the first customer of this session), clock of the exponentially distributed time starts. After this time expires, with a fixed probability, new customer of this session arrives, the clock restarts, etc. With the complementary probability, the session is considered finished. In the paper Lee et al. (2007), multi-server queueing model with a finite buffer, stationary Poisson arrival process of sessions and exponential service time distribution was considered. The stationary distribution of the system states was computed and the problem of optimal choice of the number of sessions, which can be processed in the system at the same time, is numerically solved. The problem of computation of the Laplace–Stieltjes transform of the sojourn time of a session (as a time since the session arrival instant till the moment when the session ends itself and all customers arrived within this session finish service) is solved. In the paper Kim et al. (2009), a single server queueing system with a finite buffer, more general, MAP, arrival process of sessions and phase-type (PH) distribution of service time is analyzed. In Dudin and Lee (2010), the results of Kim et al. (2009) are extended to the case of a system with an infinite buffer. Here, the steady-state analysis of the system was preceded by derivation of the stability condition of the system.

Significantly more general session-based mechanism of customers arrival was offered and analyzed in the paper Kim et al. (2012). Here, sessions arrive in the MAP and customers within each session arrive according to a terminating BMAP process. This allowed to essentially weaken strict assumptions made in Lee et al. (2007) (intervals between customers arrival in a session are independent identical exponentially distributed random variables and the number of customers in each session has the geometrical distribution). The model of session arrivals considered in Kim et al. (2012) allows to intervals between successive customers arrival be correlated, having non-identical and non-exponential distribution. Distribution of the number of customers in an arbitrary session can have quite general form. Customers can arrive in batches of a random size.

It is worth to note that the session arrivals are already considered also in context of the retrial queues. In Kim and Dudin (2012), the model with sessions retrial in the case when the session can not be admitted to the system immediately upon arrival was analyzed while in Dudin (2009) retrials of customers within an admitted session are considered.

All the mentioned results concern the systems with homogeneous sessions. Definitely, in real-life systems, the users generating the sessions may have various importance for the system and can be quite different with respect to the expected length of the session and the speed of generating the requests. So, consideration of heterogeneous sessions with different priorities implemented in the paper by the authors for discrete-time has to be made for continuous-time as well.

References

Bruneel H, Kim BG (1993) Discrete-time models for communication systems including ATM. Kluwer Academic, Boston

Chakravarthy SR (2001) The batch Markovian arrival process: a review and future work. In: Krishnamoorthy A et al (eds) Advances in probability theory and stochastic process. Notable, NJ, pp 21–49

Dudin SA (2009) The MAP/N/N retrial queueing system with time-phased batch arrivals. Probl Inf Transm 45:270–281

Dudin S, Lee MH (2010) Queues with session arrivals as models for optimizing the traffic control in telecommunication networks. Trends in telecommunications technologies. In-tech, Vukovar, pp 123–144

Heyman DP, Lucantoni D (2003) Modelling multiple IP traffic streams with rate limits. IEEE/ACM Trans Netw 11:948–958

Kim CS, Dudin S (2012) A multi-server queueing model with retrial connection arrivals as a model for optimization of the traffic control. Int J Syst Sci 43:1555–1567

Kim CS, Dudin SA, Klimenok VI (2009) The MAP/PH/1/N queue with time phased arrivals as model for traffic control in telecommunication networks. Perform Eval 66:564–579

Kim CS, Dudin A, Dudin S, Klimenok V (2012) Queueing system with batch arrival of customers in sessions. Comput Ind Eng 62:890–897

Kist AA, Lloyd-Smith B, Harris RJ (2005) A simple IP flow blocking model. Performance challenges for efficient next generation networks. In: Proceedings of 19th international teletraffic congress, pp 355–364.

Klemm A, Lindermann C, Lohmann M (2003) Modelling IP traffic using the batch Markovian arrival process. Perform Eval 54:149–173

Lee MH, Dudin S, Klimenok V (2007) Queueing model with times-phased batch arrivals. In: Managing traffic performance in converged networks the interplay of convergent and divergent forces. Proceedings of the 20th international teletraffic congress. Lecture notes in computer science, vol 4516, pp 716–730.

Lucantoni DM (1991) New results on the single server queue with a batch Markovian arrival process. Commun Stat Stoch Models 7:1–46