Analysis of the transient delay in a discrete-time buffer with batch arrivals

Joris Walraevens, Dieter Fiems and Herwig Bruneel
Department of Telecommunications and Information Processing (IR07)
Ghent University - UGent
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.
E-mail: {jw,df,hb}@telin.UGent.be

Abstract

We perform a discrete-time analysis of the delay of customers in a buffer with batch arrivals. We characterize the delay of the k-th customer that enters a FIFO buffer. The numbers of arrivals per slot are independent and identically distributed stochastic variables. Since the arrivals come in batches, the delays of the subsequent customers do not constitute a Markov-chain, which complicates the analysis. By using generating functions and the supplementary variable technique, moments of the delay of the k-th customer are calculated.

1 Introduction

The delay experienced by customers is one of the most important characteristics of a queueing system. While other characteristics (the queue content, the loss ratio, . . .) are especially interesting from the point of view of the system, the delay is arguably the most important performance measure from the users' perspective. This is especially the case for multimedia applications in current telecommunication networks, since timely delivery of their packets is of the utmost importance for these applications.

It is therefore obviously interesting to analyze the delay in queueing systems. Subsequently, there is indeed a vast literature dealing with the *steady-state* analysis of the delay in a broad range of queueing systems. However, only few research efforts have addressed the derivation of transient delay measures [3–6].

The mentioned papers all analyze *single arrival* systems. Their studies are thus not (directly) applicable to queueing systems with *batch arrivals*. In this paper, we study the transient delay in a queueing system with batch arrivals. The batch sizes are independent and identically distributed (i.i.d.) stochastic variables with an arbitrary distribution. This batch nature of the arrivals complicates the analysis since one has to keep track of the ordinal number of the customers in their arrival batch. The delays of subsequently arriving customers in a $G^{GI}/G/1$ queue do *not* constitute a Markov chain, while the corresponding delays in the $G/G/1$ queue do. Note that the aim of the present paper is to present an analytic technique to derive exact time-dependent delay measures of discrete-time buffer systems with (generally distributed) batch arrivals. We have therefore chosen to analyze the discrete-time $M^{GI}/D/1$ queue, because it is general enough to capture systems with batch arrivals, and, on the other hand, simple enough to allow a clear presentation of our method of analysis.

The remainder is organized as follows. The queueing model is described in more detail in the following section. The transient delay analysis is summarized in section 3. Several examples are discussed in section 4, and some final comments are given in section 5.

2 Queueing Model

We consider a discrete-time single-server FIFO queueing system with infinite buffer space. The number of arrivals during slot l is denoted by a_l. The a_l ($l = 0, 1, \ldots$) are i.i.d. stochastic variables. We use the notations $a(n)$ and $A(z)$ to indicate their common probability mass function (pmf) and probability generating function (pgf) respectively, i.e., $a(n) \triangleq \text{Prob}[a_l = n]$, $n \geq 0$ and $A(z) \triangleq \text{E}[z^{a_l}] = \sum_{n=0}^{\infty} a(n)z^n$. We make no assumptions on the specific arrival instants within a slot. However, it is implicitly assumed that customers arrive in a certain order within their arrival slot. This is necessary to make "the k-th arriving customer" a valid term.

3 Analysis

The delay d_k of the k-th arriving customer is defined as the number of slots between the end of the customer's arrival slot and the end of his departure slot (thus excluding his arrival slot and including his departure slot). Because of the general distribution of the arrival batch sizes, the series $\{d_k, k \geq 1\}$ does not constitute a Markov chain. Therefore we introduce additional stochastic variables r_k ($k \geq 1$) defined as the ordinal number of the k-th arriving customer in his arrival batch. With this definition, $\{(r_k, d_k), k \geq 1\}$ is easily seen to constitute a Markov chain. Let t_{k-1} denote the interarrival time (expressed in a number of slots) between customers $k-1$ and k. The following equations are then established:

$$r_k = r_{k-1}\mathbf{1}_{\mathbf{t_{k-1}=0}} + 1, \quad d_k = [d_{k-1} - t_{k-1}]^+ + 1. \quad (1)$$

Here, $\mathbf{1_X}$ denotes the indicator function of the event X.

The next step in the analysis is the introduction of generating functions and the solution of the model in the z-domain. The ultimate goal is to find an expression for the generating function of the sequence of the probability generating functions $\left\{\mathrm{E}\left[z^{d_k}\right], k \geq 1\right\}$ with respect to the discrete time parameter k:

$$D(x,z) \triangleq \sum_{k=1}^{\infty} \mathrm{E}\left[z^{d_k}\right] x^k. \quad (2)$$

Therefore we first condition on the value of r_k and express the partial pgf $D_k^{(i)}(z)$ of d_k given $r_k = i$ ($i > 1$) as funtions of $D_k^{(1)}(z)$ by use of (1) for $t_{k-1} = 0$. This expression permits us to establish a relationship between $D(x,z)$ and $D^{(1)}(x,z)$, defined as $D^{(1)}(x,z) \triangleq \sum_{k=1}^{\infty} D_k^{(1)}(z)x^k$. We then first establish a functional equation for $D^{(1)}(x,z)$ which is then transformed to a functional equation for $D(x,z)$. The functional equation for $D^{(1)}(x,z)$ is retrieved by the use of (1) for $t_{k-1} = 0$. The final expression for the functional equation for $D(x,z)$ reads

$$\begin{aligned} D(x,z) = &\frac{A(xz)-1}{(1-A(0))(xz-1)(z-A(xz))} \\ &\times \Bigg((z-A(0))x\mathrm{E}\left[z^{d_1}\right] + D(x,A(0)) \\ &\times \frac{z(A(A(0)x)-A(0))(A(0)x-1)(z-1)}{A(0)(A(A(0)x)-1)}\Bigg). \quad (3) \end{aligned}$$

In order to determine the boundary function $D(x,A(0))$, we use the analyticity of $D(x,z)$ in the unit disk $\{z : |z| < 1\}$, for all values of x in $\{x : |x| < 1\}$ and Rouché's theorem. $V(x)$, the unique zero of $z - A(xz)$ inside the unit disk of the z-plane for all $|x| < 1$, plays hereby a crucial role. We finally find

$$\begin{aligned} D(x,z) = &\frac{x(A(xz)-1)}{(1-A(0))(xz-1)(z-A(xz))}\Bigg((z-A(0)) \\ &\times \mathrm{E}\left[z^{d_k}\right] - \frac{z(z-1)(V(x)-A(0))}{V(x)(V(x)-1)}\mathrm{E}\left[V(x)^{d_k}\right]\Bigg). \end{aligned} \quad (4)$$

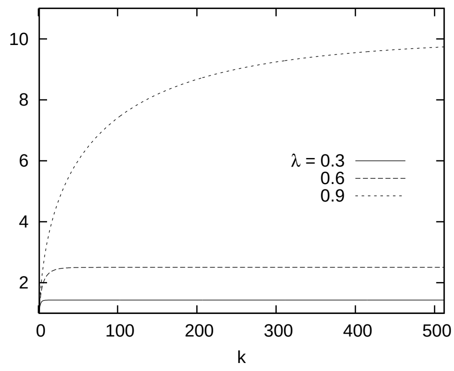

Figure 1. Transient mean delay for customers arriving in geometrically sized batches with mean size 0.3, 0.6 and 0.9 respectively.

It then remains to calculate $V(x)$. In general, it is not possible to find an explicit expression for $V(x)$. However, we can calculate this function in each argument numerically as follows: starting from $i = 0$, calculate $V_{i+1}(x)$ as $V_{i+1}(x) = A(xV_i(x))$. It can be proved that $\lim_{i\to\infty} V_i(x) = V(x)$ [2], regardless of the starting point $V_0(x)$ inside the complex unit disk. From expression (4), various performance measures can be calculated. In particular, let $\bar{d}_k$ denote the mean delay of the k-th customer. The generating function of the sequence $\{\bar{d}_k\}$ is given by the partial derivative to z in 1 of $D(x,z)$ and an expression for this generating function can be thus be found from (4).

Finally, the transforms have to be numerically inverted. Most mathematical software packages provide procedures for numerical inversion of transform functions. For more details on a specific procedure with known error bounds, we refer to [1]. However, the calculated error bounds are only practical (i.e. small enough) for stable systems, i.e., where the mean delay of the k-th customer goes to a finite steady-state value for $k \to \infty$.

4 Numerical example

In this section, we show a small example. In Figure 1, we show the mean delay of customer k as a function of k. The batch sizes are geometrically distributed random variables with mean λ and the buffer is empty at the beginning. It can be seen from this figure (and other examples) that the mean transient delays tend to the steady-state values rather quickly for small λ while the convergence is much slower for larger λ.

5 Conclusions

In this paper, we have analyzed the transient delay in a discrete-time $M^{GI}/D/1$ queue by using the supplementary variable technique and by further making extensive use of generating functions. A functional equation was obtained and solved for the transform of the probability generating functions of the delays of the customers. From this transform various performance measures can be calculated, most notably the mean transient delay.

Acknowledgment. The first two authors are Postdoctoral Fellows with the Fund for Scientific Research, Flanders (F.W.O.-Vlaanderen), Belgium.

References

[1] J. Abate and W. Whitt. Numerical inversion of probability generating functions. *Operations Research Letters*, 12(4):245–251, 1992.

[2] J. Abate and W. Whitt. Solving probability transform functional equations for numerical inversion. *Operations Research Letters*, 12(5):275–281, 1992.

[3] D. Bertsimas and G. Mourtizinou. Transient laws of non-stationary queueing systems and their applications. *Queueing Systems*, 25(1-4):115–155, 1997.

[4] T. Hofkens, K. Spaey, and C. Blondia. Transient analysis of the D-BMAP/G/1 queue with an application to the dimensioning of a playout buffer for VBR video. *Lecture Notes in Computer Science*, 3042:1338–1343, 2004.

[5] C. Wang. On the transient delays of M/G/1 queues. *Journal of Applied Probability*, 36(3):882–893, 1999.

[6] C. Wang. An identity of the GI/G/1 transient delay and its applications. *Probability in the Engineering and Informational Sciences*, 16(1):47–66, 2002.