Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics, Istanbul, Turkey, May 27-29, 2006 (pp186-191)

Buffer Overflow Period in a Constant Service Rate Queue

ANDRZEJ CHYDZINSKI*
Institute of Computer Sciences,
Silesian University of Technology
Akademicka 16, 44-100 Gliwice
POLAND

Abstract: – When we speak about the buffer overflow period we mean the time interval in which the queue size is equal to the capacity of the buffer. Thus packets or cells arriving during this period are lost. The distribution of the length of this interval is recently gaining attention as a very informative characteristic of the queueing performance. In this paper the numerous results related to the distribution of the buffer overflow period in a constant service time queue are presented. In particular, they include formulas for distribution functions, probability density functions, expected values, limiting distributions for a large buffer. The queueing models supporting single and batch arrivals are considered.

Key-Words: – performance evaluation, queueing systems, buffer overflow

1 Introduction

The classical approach for evaluation of queueing performance of traffic buffering in network switches is based on such characteristics as steady-state (or transient) queue size distribution, actual (or virtual) waiting time distribution, cell (or packet) loss ratio. Recently one may observe a growing popularity of other performance measures. Some authors indicate that, instead of widely used cell loss ratio, the transient time until the first buffer overflow occurs can be more meaningful. Detailed arguments are presented, for instance, in [13] and they are based on widely-known properties of observed traces of networks' traffic. The time of the first buffer overflow is also called the first passage time and the recent account of the computational techniques for finding its distribution may be found by the reader in [3].

Another very informative characteristic is the duration of the buffer overflow period. The buffer overflow period is the period in which the buffer is full and arrivals are not accepted. It also can be defined as the remaining service time upon reaching a full buffer state. The importance of the distribution of the buffer overflow period is connected with the fact, that its form is responsible for the probabilistic structure of losses caused by the buffer overflow. In other words, two systems with the same loss ratio can behave in a very different way if they have different distributions of the buffer overflow period. In one of them we may observe rather single losses, while in the other the losses may have tendency to occur in groups.

In order to distinguish the difference between such queueing systems we use the n-consecutive cell loss probability ($n-$CCL prob.). It is said to be the probability, that during one busy cycle, at least once a group of n consecutive arrivals are all lost owing to buffer overflow. Knowing the distribution of the buffer overflow period, we are able to derive $n-$CCL probabilities (see [4], formulas (26)-(28)).

The duration of the buffer overflow period is not only an interesting characteristic in its own right but also appears as a problem in the diffusion approximation of finite capacity systems (see [12]) or in analysis of threshold-based queues (see [6]).

*This material is based upon work supported by the Polish Ministry of Scientific Research and Information Technology under Grant No. 3 T11C 014 26.

In this paper a collection of results on the distribution of the buffer overflow period in a constant service time queue is presented. The following model of the queue is considered: Cells (packets, customers) arrive according to a Poisson process with intensity ν and they are served individually with the constant service time equal to d. The capacity of the buffer is finite and equal to b (including service position). In Kendall's notation such a system is denoted by $M/D/1/b$. In addition, in section 4 a batch arrival system ($M^X/D/1/b$) is considered. Namely, we allow arrivals of groups of cells where sizes of consecutive groups are independent, identically distributed with discrete distribution $\{a_0, a_1, a_2, \ldots\}$, $\sum_{i=0}^{\infty} a_i = 1$.

The system with the constant service time is worth studying at least for two reasons. Firstly, the constant service time in the meaning of transmission time is one of the most important types of service time from the practical point of view. Secondly, its special form causes significant simplification in the analysis and therefore the results are less complex in numerical computations and easy to apply. Detailed calculations and proofs are based on a well-established methodology and the interested reader can easily reconstruct them using the remarks and references given below the theorems.

For the previous results connected with the duration of the buffer overflow period (or, equivalently, remaining service time) we refer the reader to [2, 4, 7, 8, 9, 10, 11, 12]. All of them, beside [12], are devoted to the general case, in which the form of the service time distribution is not further specified. In particular, the equilibrium distributions of remaining and past service times upon reaching a target level in the $M/G/1$ queueing system are presented in [2]. In [10], chapter II.5.10, studies of a similar subject, namely the remaining interarrival time in a $G/M/1$ queue are presented. Some of the properties of the average remaining service time in the $G/G/1$ model are shown in [11]. In paper [12], approximate, based on heuristics formula for the mean duration of the buffer overflow period in the case of constant service time and a Poisson input stream is obtained. The limiting distribution (as $b \to \infty$) of the remaining service time distribution in an $M/G/1/b$ queue is determined in [4]. In [7], an explicit form of the buffer overflow period distribution in the $M/G/1/b$ model is shown. Finally, the analysis of the buffer overflow period in a batch arrival queue with a general service time is carried out in [8, 9]. In particular, formulas for the distribution of this period and their asymptotic behaviour (as buffer size goes to infinity) are presented.

2 Definitions and notation

The buffer overflow period is formally defined in the following way. Let $X(t)$ denote the queue size at the moment, t of the system. Let the initial queue size be $X(0) = n$ where $0 \leq n < b$. By $\tau^+(n, b)$ we denote the time of the first buffer overflow (or first passage time). Formally $\tau^+(n, b) = \inf\{t > 0 : X(t) = b\}$. Let $\zeta(n, b)$ stands for the first departure moment after $\tau^+(n, b)$. The *buffer overflow period* is defined to be $\beta(n, b) = \zeta(n, b) - \tau^+(n, b)$.

Usually we pay a special attention to two initial queue sizes. For $n = 0$ the distribution of the buffer overflow period is called the *first hit distribution*, while for $n = b - 1$ – the *subsequent hit distribution*. In queueing systems with a Poisson input stream only the duration of the first buffer overflow period depends on the initial queue length n, every other has the initial queue length equal to $b - 1$. Therefore we may think of the first hit distribution as a transient characteristic while the subsequent hit distribution as a stationary one.

Throughout the article the following notation will be used:

$\mathbf{P}(\,\cdot\,)$ - the probability

ν – the intensity of the Poisson input stream

d – the duration of the service of one cell (constant)

b – the capacity of the buffer

$H_{n,b}(z) = \mathbf{P}(\beta(n, b) < z)$ – the distribution function of the remaining service time

$H_b(z) = \mathbf{P}(\beta(b-1, b) < z)$ – the subsequent hit distribution function

$H(z) = \lim_{b\to\infty} H_b(z)$ – the subsequent hit distribution function for a large buffer

$\overline{H}(z) = \lim_{b\to\infty} H_{0,b}(z)$ – the first hit distribution function for a large buffer

$h_{n,b}(z) = H'_{n,b}(z)$ – the probability density function of the remaining service time

$h_b(z) = H_b'(z)$ – the subsequent hit probability density function

$h(z) = H'(z)$ – the subsequent hit density function for a large buffer

$\overline{h}(z) = \overline{H}'(z)$ – the first hit density function for a large buffer

M – the expected value for $H(z)$,

m – the expected value for $\overline{H}(z)$

$$I(x \le y) = \begin{cases} 1 & \text{if} \quad x \le y, \\ 0 & \text{if} \quad x > y. \end{cases}$$

$\delta_{i,j}$ - the Kronecker symbol ($\delta_{i,j} = 1$ if $i = j$ and 0 otherwise)

3 Results for a single-arrival input stream

In this section we assume that cells arrive singly. Thus the offered load (traffic intensity) is just

$$\rho = \nu d.$$

Theorem 1. *For the $M/D/1/b$ queue it holds true that:*

$$H_b(z) = 1 - \frac{\sum_{k=1}^{b-1} A_k(z)(R_{b-k} - R_{b-1-k})}{R_{b-1} - R_{b-2} + \delta_{b,1} - \delta_{b,2}}, \qquad (1)$$

$$H_{n,b}(z) = 1 + \sum_{k=1}^{b-n} A_k(z) R_{b-n-k} - (1 - H_b(z))(R_{b-n-1} + \delta_{b-n,1}), \qquad (2)$$

$$h_b(z) = -\frac{\sum_{k=1}^{b-1} a_k(z)(R_{b-k} - R_{b-1-k})}{R_{b-1} - R_{b-2} + \delta_{b,1} - \delta_{b,2}}, \qquad (3)$$

$$h_{n,b}(z) = \sum_{k=1}^{b-n} a_k(z) R_{b-n-k} + h_b(z)(R_{b-n-1} + \delta_{b-n,1}), \qquad (4)$$

where

$$A_k(z) = I(z \le d)\nu \int_0^{d-z} \frac{e^{-\nu u}(\nu u)^{k-1}}{(k-1)!} du, \qquad (5)$$

$$a_k(z) = I(z \le d)\nu \frac{-((d-z)\nu)^{k-1} e^{-(d-z)\nu}}{(k-1)!}, \qquad (6)$$

and

$$R_0 = 0,$$

$$R_{k+1} = e^{\nu d}\Big(R_k - e^{-\nu d} \sum_{n=0}^{k} \frac{(\nu d)^{n+1} R_{k-n}}{(n+1)!} + \delta_{0,k}\Big), \quad k \ge 0. \qquad (7)$$

The proof of (1) and (2) can be obtained using the same mathod as in the proof of Theorem 3 from [7]. (3), (4) and (6) were obtained by calculating derivatives of (1), (2) and (5), respectively.

Theorem 2. *For the $M/D/1/b$ queue with $\rho \ne 1$ it holds that:*

$$H(z) = 1 - I(z \le d) \frac{e^{(x_0-1)(d-z)\nu} - 1}{x_0 - 1}, \qquad (8)$$

$$\overline{H}(z) = \frac{H(z) - \rho + I(z \le d)\nu(d-z)}{1-\rho}, \qquad (9)$$

$$h(z) = I(z \le d)\nu e^{(x_0-1)(d-z)\nu}, \qquad (10)$$

$$\overline{h}(z) = \frac{h(z) - I(z \le d)\nu}{1-\rho}, \qquad (11)$$

$$M = \frac{e^{(x_0-1)\nu d} - \nu d(x_0-1) - 1}{(x_0-1)^2 \nu}, \qquad (12)$$

$$m = \frac{M - \rho d/2}{1-\rho}. \qquad (13)$$

where x_0 is a positive solution of the equation

$$e^{\nu(x-1)d} = x.$$

Methodology for obtaining $h(z)$ and $\overline{h}(z)$ is given in [4], where the general case is investigated. Then, by integrating (10) and (11), it is easy to get (8) and (9). (12), (13) follow easily from (8) and (9).

4 Results for a batch-arrival input stream

In this section we extend the queueing model by allowing batch arrivals. Precisely, instead of single arrivals we assume arrivals of groups of cells, where sizes of groups are independent and identically distributed with discrete distribution $\{a_0, a_1, a_2, \ldots\}$, $\sum_{i=0}^{\infty} a_i = 1$. Number a_i denotes the probability of the event that arriving batch has size i. Note, that if $a_0 > 0$, then in fact some arrivals are virtual and they cause only an extension of the interarrival time.

This makes the system more flexible for modeling the input stream. As previously, cells are served individually and service time for one cell is d. For a recent methodology for finding basic characteristics (queue size, waiting time) of such a system we refer the reader to [5].

By $a(x)$ we denote here the generating function for distribution $\{a_i\}$, namely

$$a(x) = \sum_{i=0}^{\infty} x^i a_i.$$

The offered load of the system is

$$\rho = \nu d a'(1).$$

Theorem 3. *For the $M^X/D/1/b$ queue it holds true that:*

$$H_b(z) = 1 - \frac{C_b(z) - I(z \le d)(1 - \sum_{k=0}^{b-1} a_k)}{\sum_{k=1}^{b} R_{k-1}a_{b-k} + a_{b-1} - R_{b-1} - \delta_{b,1}}, \quad (14)$$

$$H_{n,b}(z) = 1 + \sum_{k=1}^{b-n} A_k(z) R_{b-n-k} - (1 - H_b(z))(R_{b-n-1} + \delta_{b-n,1}), \quad (15)$$

$$h_b(z) = \frac{-c_b(z)}{\sum_{k=1}^{b} R_{k-1}a_{b-k} + a_{b-1} - R_{b-1} - \delta_{b,1}}, \quad (16)$$

$$h_{n,b}(z) = \sum_{k=1}^{b-n} a_k(z) R_{b-n-k} + h_b(z)(R_{b-n-1} + \delta_{b-n,1}), \quad (17)$$

where the sequences R_k, $C_k(z)$, $c_k(z)$, $A_k(z)$ and $a_k(z)$ have generating functions:

$$R(x) = \sum_{k=0}^{\infty} x^k R_k = \frac{x}{e^{-\nu d(1-a(x))} - x}, \quad (18)$$

$$C(x,z) = \sum_{k=1}^{\infty} x^k C_k(z) = I(z \le d)\frac{x^2(a(x)-1)(e^{-\nu(d-z)(1-a(x))} - 1)}{(x-1)(e^{-\nu d(1-a(x))} - x)}, \quad (19)$$

$$c(x,z) = \sum_{k=1}^{\infty} x^k c_k(z) = -I(z \le d)\frac{\nu x^2(a(x)-1)^2 e^{-\nu(d-z)(1-a(x))}}{(x-1)(e^{-\nu d(1-a(x))} - x)}, \quad (20)$$

$$A(x,z) = \sum_{k=1}^{\infty} x^k A_k(z) = I(z \le d)\frac{x}{x-1}(e^{-\nu(d-z)(1-a(x))} - 1), \quad (21)$$

$$a(x,z) = \sum_{k=1}^{\infty} x^k a_k(z) = I(z \le d)\frac{\nu x(1-a(x))}{x-1}(e^{-\nu(d-z)(1-a(x))}). \quad (22)$$

Formulas for $H_b(z)$, $H_{n,b}(z)$ can be proven using the methodology given in [9]. Densities (16), (17) were obtained by differentiation. We have to be careful when using them, as they are sometimes defective (integrable to $c < 1$). This is due to the fact that distributions $H_b(z)$, $H_{n,b}(z)$ may have a discrete component, namely an atom of probability mass at point $z = d$. There is a simple explanation of this phenomena: one, large enough, batch can overflow the buffer in one step starting from the empty queue and buffer overflow period has length d in this case. As regards the numerical computation aspect of (14), (15), we do not have explicit formulas for coefficients R_k, $C_k(z)$ and $A_k(z)$. To overcome this problem, we can use one of the procedures for the numerical generating function inversion [1].

Theorem 4. *For the $M^X/D/1/b$ queue with $\rho \ne 1$ it holds that:*

$$H(z) = 1 - I(z \le d)\frac{e^{(a(x_0)-1)(d-z)\nu} - 1}{x_0 - 1}, \quad (23)$$

$$\overline{H}(z) = \frac{H(z) - \rho + I(z \le d)a'(1)\nu(d-z)}{1-\rho}, \quad (24)$$

$$h(z) = I(z \le d)\frac{(a(x_0)-1)\nu e^{(a(x_0)-1)(d-z)\nu}}{x_0 - 1}, \quad (25)$$

$$\overline{h}(z) = \frac{h(z) - I(z \le d)a'(1)\nu}{1-\rho}, \quad (26)$$

$$M = \frac{e^{(a(x_0)-1)\nu d} - \nu d(a(x_0)-1) - 1}{(x_0-1)(a(x_0)-1)\nu}, \quad (27)$$

$$m = \frac{M - \rho d/2}{1-\rho}. \quad (28)$$

where x_0 is a positive solution of the equation

$$e^{\nu(a(x)-1)d} = x.$$

Methodologies for obtaining $H(z)$ and $\overline{H}(z)$ are presented in [9] and [8], respectively. Then, obtaining densities and expected values is straightforward.

5 Numerical examples

Now we can easily obtain a great variety of numerical examples depending on parameters of the systems (ν, d, b, $\{a_i\}$) and on the type of characteristic we want to analyse. Figs. 1 - 5 present sample results for single-arrival input stream, while Figs. 6, 7 present sample results for a batch-arrival input stream with the following batch size distribution:

$$a_i = 0.1, \quad i = 1, \ldots, 10.$$

In particular, in Fig. 1 the dependence of the subsequent hit density, $h_b(z)$, on b is presented. Fig. 2 shows the dependence of the subsequent hit density, $h_b(z)$, on ν. In Fig. 3 the dependence of subsequent hit density, $h_b(z)$, on d is displayed. Figures 4, 5 present the comparison of the first and the subsequent densities for a large buffer ($h(z)$ vs $\overline{h}(z)$). Figs. 6, 7 show the average duration of the buffer overflow period for a large buffer as a function of ν and d, respectively.

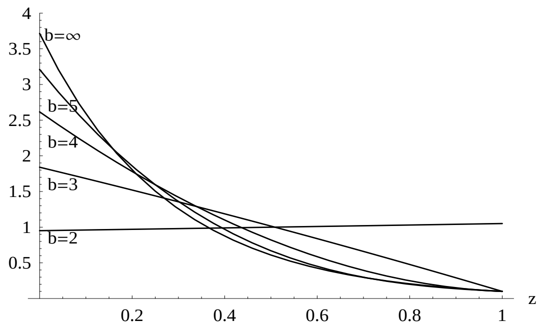

FIGURE 1. The dependence of subsequent hit density, $h_b(z)$, on b. Single-arrival input stream, $d = 1$, $\nu = 0.1$.

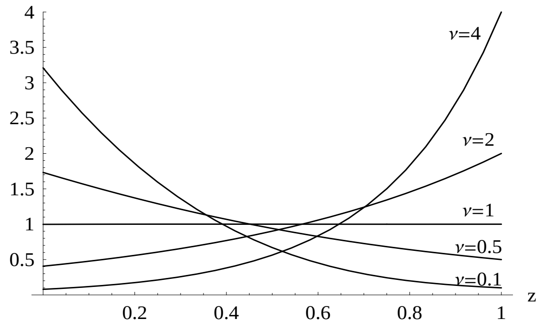

FIGURE 2. The dependence of subsequent hit density, $h_b(z)$, on ν. Single-arrival input stream, $d = 1$, $b = 5$.

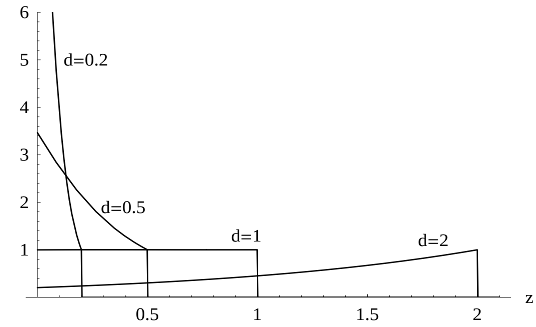

FIGURE 3. The dependence of subsequent hit density, $h_b(z)$, on d. Single-arrival input stream, $\nu = 1$, $b = 5$.

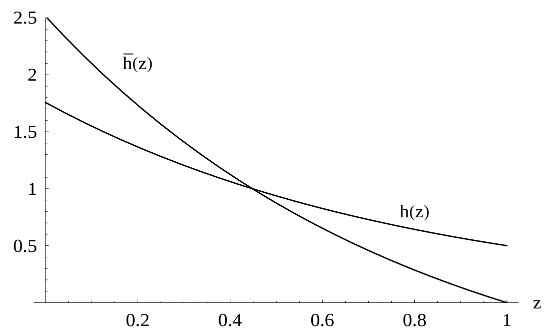

FIGURE 4. First and subsequent densities for large buffer ($h(z)$ vs $\overline{h}(z)$). Single-arrival input stream, $d = 1$, $\nu = \rho = 0.5$.

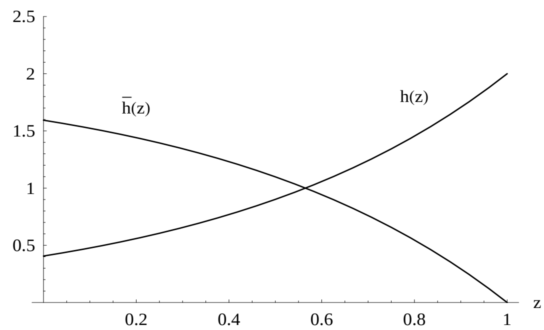

FIGURE 5. First and subsequent densities for large buffer ($h(z)$ vs $\overline{h}(z)$). Single-arrival input stream, $d = 1$, $\nu = \rho = 2$.

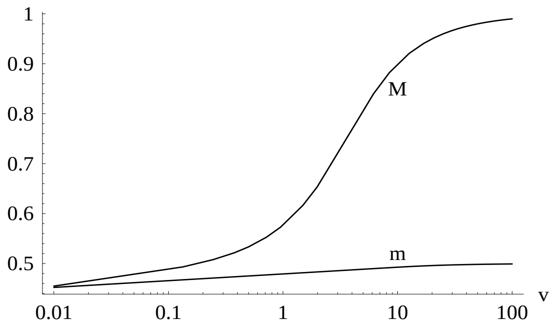

FIGURE 6. The average duration of the buffer overflow period for a large buffer as a function of ν. Batch-arrival input stream, $d = 1$, $a_i = 0.1$ for $i = 1, \ldots, 10$.

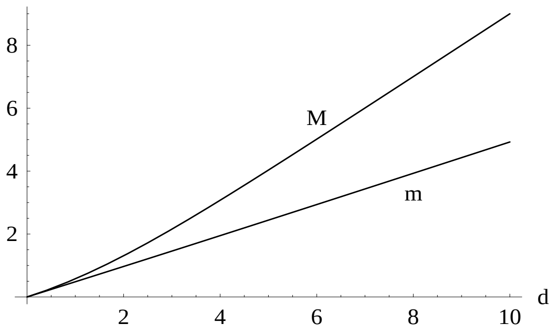

FIGURE 7. The average duration of the buffer overflow period for a large buffer as a function of d. Batch-arrival input stream, $\nu = 1$, $a_i = 0.1$ for $i = 1, \ldots, 10$.

6 Conclusions

In the paper formulas for ten basic characteristics ($H_{n,b}$, H_b, H, $\overline{H}$, $h_{n,b}$, h_b, h, $\overline{h}$, M and m) of the distribution of the buffer overflow period are presented, both for systems with single and batch arrivals. The duration of the buffer overflow period is an interesting and informative characteristic of the queueing performance.

The numerical examples indicate that, even for such a simple service time distribution as a deterministic one, there exists a surprising diversity of shapes of the distribution of the buffer overflow period. Due to this fact, the consecutive losses may have significantly different statistical structure depending on system parameters. Therefore, the buffer overflow period can seriously influence the performance of the system.

References

[1] J. Abate, G. L. Choudhury and W. Whitt. An introduction to numerical transform inversion and its application to probability models. Chapter in Computational Probability, pp. 257-323, W. Grassman (ed.), Kluwer, Boston, 1999.

[2] S. Asmussen. Equilibrium properties of the M/G/1 queue. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 58:267-281, 1981.

[3] S. Asmussen, M. Jobmann and H.-P. Schwefel. Exact Buffer Overflow Calculations for Queues via Martingales. Queueing Systems 42(1): 63-90, 2002.

[4] P. T. de Boer, V.F. Nicola and J.C.W. van Ommeren. The Remaining Service Time Upon Reaching a High Level in M/G/1 Queues. Queueing Systems, 39:55-78, 2001.

[5] M. Bratiychuk and B. Borowska. Explicit formulae and convergence rate for the system $M^{\alpha}/G/1/N$ as $N \to \infty$. Stochastic Models, 18(1), pp. 71-84, 2002.

[6] M. Bratiychuk and A. Chydzinski. On the ergodic distribution of oscillating queueing systems. Journal of Applied Mathematics and Stochastic Analysis. 16:3, 2003.

[7] A. Chydzinski. On the remaining service time upon reaching a target level in M/G/1 queues. Queueing Systems, vol. 47, issue 1/2, pp. 71-80, 2004.

[8] A. Chydzinski, On the Transient Remaining Service Time in a Large Buffer Queue, WSEAS Transactions on Communications. Issue 4, Vol. 3, pp. 958-962, 2004.

[9] A. Chydzinski. Duration of the buffer overflow period in a batch arrival queue. Performance Evaluation, 63:4-5, pp. 493-508, 2006.

[10] J. W. Cohen. The single server queue. North-Holland, Amsterdam, 2nd edition, 1982.

[11] D. Fakinos. The expected remaining service time in a single server queue. Operations Research, 30:1014-1018, 1982.

[12] E. Gelenbe, X. Mang and R. Önvural. Diffusion based statistical call admission control in ATM. Performance Evaluation, 27&28:411-436, 1996.

[13] H.-P. Schwefel, L. Lipsky and M. Jobmann. On the necessity of transient performance analysis in telecommunication systems, in: *Teletraffic Engineering in the Internet Era*, eds. J.M. de Souza, N.L.S. da Fonseca and E.A.S. Silva (Elsevier, Amsterdam, 2001).