Faculteit der Economische Wetenschappen en Econometrie

Serie Research Memoranda

Approximations for the Overflow Probability in Finite-Buffer Queues

Henk C. Tijms

Research Memorandum 1991-59
August 1991

vrije Universiteit amsterdam

APPROXIMATIONS FOR THE OVERFLOW PROBABILITY IN FINITE-BUFFER QUEUES.

Henk C. Tijms, Vrije Universiteit, Amsterdam

Abstract: This paper presents and motivates a new heuristic for the overflow probability in finite-buffer queues.

Zusammenfassung: Diese Arbeit gibt eine Motivierung für eine neue Approximation für die Verlustwahrscheinlichkeit in Wartesystemen mit beschränkter Kapazität.

1. Introduction

A practically important problem in computer and communication systems is the calculation of the overflow probability in finite-capacity queueing systems. Consider a queueing system where customers arrive and depart one at a time, and where the system capacity is N customers (including any customer in service). A customer who finds upon arrival N other customers in the system is turned away and has no further effect on the system. A common heuristic for the overflow probability (or the long-run fraction of customers that is turned away) is

$$\sum_{j=N}^{\infty} \pi_j^{(\infty)}, \tag{1}$$

where $\{\pi_j^{(\infty)}\}$ is the equilibrium distribution of the number of customers present just prior to an arrival epoch in the corresponding *infinite-capacity* model. That is, the overflow probability is approximated by the equilibrium probability that in the infinite-capacity model a customer finds upon arrival N or more other customers present. Numerical investigations in /1/ indicate that this heuristic which is commonly used in practice may perform very poorly. Therefore the following new approximation for the overflow probability was proposed in /1/:

$$\frac{(1-\rho)\sum_{j=N}^{\infty}\pi_j^{(\infty)}}{1-\rho\sum_{j=N}^{\infty}\pi_j^{(\infty)}}, \tag{2}$$

where ρ denotes the traffic intensity for the (infinite-capacity) queueing model. The more refined heuristic (2) performs much better than the crude heuristic (1). It is practically very useful for dimensioning the buffer size, including situations in which an extremely small overflow probability is required.

In this paper a motivation for the new heuristic (2) will be given. Using simple probabilistic arguments, it will be shown that the heuristic (2) is exact for both the finite-capacity, single-server queue M/G/1/N with general services and the finite capacity, multi-server queue M/M/c/N with exponential services.

2. The overflow probability

Consider the finite-capacity M/G/c/N queueing system in which customers arrive according to a Poisson process and any customer finding upon arrival $N(\geq c)$ other customers present is turned away. Denote by

λ = the average arrival rate of customers

α = the average service time per customer.

It is assumed that the traffic intensity

$$\rho = \frac{\lambda\alpha}{c} < 1.$$

Further, it is assumed in the sequel that for the multi-server case $(c>1)$ the service times have an *exponential* distribution. However, for the single-server case $(c=1)$ a general service-time distribution is allowed.

Defining a cycle as the time elapsed between two consecutive arrivals who find the system empty, the following concepts are introduced for the finite-capacity model:

T = the length of a cycle

T_j = the amount of time in a cycle during which j customers are present $(j=0,1,\ldots,N)$

N_j = the number of service completions in a cycle at which j customers are left behind $(j=0,1,\ldots,N-1)$.

Also, for the finite-capacity model, let

μ_j = the expected time from a service completion epoch at which j customers are left behind until the next service completion epoch $(j=0,1,\ldots,N-1)$.

The corresponding quantities for the infinite-capacity queueing model are denoted by T^{∞}, T_j^{∞}, N_j^{∞} and $\mu_j^{(\infty)}$.

The following relations are obvious:

$$E(T) = \mu_0 + \sum_{j=1}^{N-1} E(N_j)\, \mu_j, \tag{3}$$

$$E(T^{\infty}) = \mu_0^{(\infty)} + \sum_{j=1}^{\infty} E(N_j^{\infty})\, \mu_j^{(\infty)}. \tag{4}$$

Also, using the exponential assumptions made, it follows from the lack of memory of the exponential distribution that

$$\mu_j = \mu_j^{(\infty)} \qquad \text{for } j = 0,1,\ldots,N-1, \tag{5}$$

$$\mu_j^{(\infty)} = \frac{\alpha}{c} \qquad \text{for } j \geq c. \tag{6}$$

For the finite-capacity model, let

p_j = the long-run fraction of time that j customers are present $(j=0,1,\ldots,N)$.

Similarly, $p_j^{(\infty)}$ is defined for the infinite-capacity model. Then by a standard result from the theory of regenerative processes,

$$p_j = \frac{E(T_j)}{E(T)} \qquad \text{for } j = 0,1,\ldots,N. \tag{7}$$

and

$$p_j^{(\infty)} = \frac{E(T_j^{\infty})}{E(T^{\infty})} \quad \text{for } j = 0,1,\ldots\ . \tag{8}$$

Next, we relate $E(N_j)$ and $E(T_j)$. Clearly, the expected number of downcrossings in a cycle from j+1 to j customers equals the expected number of upcrossings in a cycle from j to j+1 customers. The first quantity is per definition equal to $E(N_j)$, while the second quantity is equal to $\lambda E(T_j)$ on basis of results in /2/ and the assumption of Poisson arrivals. Thus

$$E(N_j) = \lambda E(T_j) \qquad \text{for } j = 0,1,\ldots,N-1 \tag{9}$$

Similarly,

$$E(N_j^{(\infty)}) = \lambda E(T_j^{\infty}) \qquad \text{for } j = 0,1,\ldots\ . \tag{10}$$

We are now in a position to prove

<u>Theorem</u> 1. For both the M/G/1/N queue and the M/M/c/N queue

$$p_j = \gamma p_j^{(\infty)} \quad \text{for } j = 0,1,\ldots,N-1, \tag{11}$$

where

$$\gamma = \left\{1 - \rho \sum_{j=N}^{\infty} p_j^{(\infty)}\right\}^{-1} \tag{12}$$

Proof. First, it is observed that N_j has the same distribution as N_j^{∞} for any j=0,1,...,N-1. Hence it holds that

$$E(N_j) = E(N_j^{\infty}) \qquad \text{for } j = 0,1,\dots,N-1. \tag{13}$$

It now follows from (7)-(10) and (13) that

$$p_j = \gamma\, p_j^{(\infty)} \qquad \text{for } j = 0,1,\dots,N-1, \tag{14}$$

where the constant γ is given by

$$\gamma = \frac{E(T^{\infty})}{E(T)}. \tag{15}$$

Dividing both sides of (3) by $E(T^{\infty})$ and using the relations (5), (7)-(10) and (13) yields

$$\frac{1}{\gamma} = \frac{\mu_0^{(\infty)}}{E(T^{\infty})} + \sum_{j=1}^{N-1} \lambda p_j^{(\infty)}\, \mu_j^{(\infty)}. \tag{16}$$

Also, by dividing both sides of (2) by $E(T^{\infty})$ and using (6), (8) and (10), we obtain

$$1 = \frac{\mu_0^{(\infty)}}{E(T^{\infty})} + \sum_{j=1}^{N-1} \lambda p_j^{(\infty)}\, \mu_j^{(\infty)} + \frac{\lambda\alpha}{c} \sum_{j=N}^{\infty} p_j^{(\infty)} \tag{17}$$

Together (16) and (17) imply the desired result.

From Theorem 1 and the relation $\sum_{j=0}^{N} p_j = 1$, it follows that

$$p_N = \frac{(1-\rho)\sum_{j=N}^{\infty} p_j^{(\infty)}}{1-\rho \sum_{j=N}^{\infty} p_j^{(\infty)}}\,.$$

Next, using the property Poisson arrivals see time averages (see /2/), it follows for the finite-capacity model that the overflow probability equals p_N. Also, by this same property, the customer-average probabilities $\pi_j^{(\infty)}$ are equal to the time-average probabilities $p_j^{(\infty)}$ for all j. This proves the final theorem:

Theorem 2. For both the M/G/1/N queue and the M/M/c/N queue, the overflow probability equals

$$\frac{(1-\rho)\ \sum_{j=N}^{\infty}\pi_j^{(\infty)}}{1-\rho\ \ \sum_{j=N}^{\infty}\pi_j^{(\infty)}},$$

where for the infinite capacity model $\pi_j^{(\infty)}$ denotes the long-run fraction of customers who find upon arrival j other customers present.

Literature:

/1/ Tijms, H.C.
Heuristics for finite-buffer queues.
Probability in the Engineering and Informational Sciences 6 (1992), to appear.

/2/ Wolff, R.W.
Poisson arrivals see time averages.
Operations Research 30, 223-231 (1982).

1991-1	N.M. van Dijk	On the Effect of Small Loss Probabilities in Input/Output Transmission Delay Systems
1991-2	N.M. van Dijk	Letters to the Editor: On a Simple Proof of Uniformization for Continious and Discrete-State Continious-Time Markov Chains
1991-3	N.M. van Dijk P.G. Taylor	An Error Bound for Approximating Discrete Time Servicing by a Processor Sharing Modification
1991-4	W. Henderson C.E.M. Pearce P.G. Taylor N.M. van Dijk	Insensitivity in Discrete Time Generalized Semi-Markov Processes
1991-5	N.M. van Dijk	On Error Bound Analysis for Transient Continuous-Time Markov Reward Structures
1991-6	N.M. van Dijk	On Uniformization for Nonhomogeneous Markov Chains
1991-7	N.M. van Dijk	Product Forms for Metropolitan Area Networks
1991-8	N.M. van Dijk	A Product Form Extension for Discrete-Time Communication Protocols
1991-9	N.M. van Dijk	A Note on Monotonicity in Multicasting
1991-10	N.M. van Dijk	An Exact Solution for a Finite Slotted Server Model
1991-11	N.M. van Dijk	On Product Form Approximations for Communication Networks with Losses: Error Bounds
1991-12	N.M. van Dijk	Simple Performability Bounds for Communication Networks
1991-13	N.M. van Dijk	Product Forms for Queueing Networks with Limited Clusters
1991-14	F.A.G. den Butter	Technische Ontwikkeling, Groei en Arbeidsproduktiviteit
1991-15	J.C.J.M. van den Bergh, P. Nijkamp	Operationalizing Sustainable Development: Dynamic Economic-Ecological Models
1991-16	J.C.J.M. van den Bergh	Sustainable Economic Development: An Overview
1991-17	J. Barendregt	Het mededingingsbeleid in Nederland: Konjunktuurgevoeligheid en effektiviteit
1991-18	B. Hanzon	On the Closure of Several Sets of ARMA and Linear State Space Models with a given Structure
1991-19	S. Eijffinger A. van Rixtel	The Japanese Financial System and Monetary Policy: a Descriptive Review
1991-20	L.J.G. van Wissen F. Bonnerman	A Dynamic Model of Simultaneous Migration and Labour Market Behaviour
1991-21	J.M. Sneek	On the Approximation of the Durbin-Watson Statistic in O(n) Operations
1991-22	J.M. Sneek	Approximating the Distribution of Sample Autocorrelations of