University of Würzburg
Institute of Computer Science
Research Report Series

A Numerical Analysis of the $M/D^b/N$ Queueing System

Mathias Dümmler and Norbert Vicari

Report No. 235 July 1999

University of Würzburg
Department of Computer Science
Am Hubland, D-97074 Würzburg, Germany
{*duemmler*|*vicari*}*@informatik.uni-wuerzburg.de*

Presented at: 1998 INFORMS national meeting in Seattle, WA, USA

A Numerical Analysis of the $M/D^b/N$ Queueing System

Mathias Dümmler and Norbert Vicari
University of Würzburg
Department of Computer Science
Am Hubland, D-97074 Würzburg, Germany
{*duemmler*|*vicari*}*@informatik.uni-wuerzburg.de*

Abstract

In this paper, we provide a numerical analysis of the $M/D^b/N - \infty$ queueing system. This system consists of an infinite buffer and N identical batch servers with batch capacity b. It serves as a model for various types of workstations in the manufacturing context. Using an approach presented by Crommelin, we derive the distribution of the number of lots waiting in queue at arbitrary time instants and the distribution of the waiting time of lots. The results are of an exact nature and the computation is of low complexity concerning run time and memory usage.

1 Introduction

Batch servers or *bulk servers* can be found in many areas of manufacturing. For example, in the manufacturing of semiconductors, lots of wafers are usually grouped together into batches at workstations like oxidation ovens or chemical deposition [1].

Furthermore, identical servers are commonly grouped together as workstations. All servers at a workstation are fed by a single queue. In this way, the idle times of the servers are minimized.

Analytical approaches for single batch servers are well known. Neuts [2] presented an analytical study of batch service systems that operate according to a minimum batch size rule. Chaudhry and Templeton [3] provide an extensive discussion of bulk service systems and their analysis. In [4] and [5], batch service systems in push and pull manufacturing environments are analyzed using embedded Markov chain techniques.

The method of analysis proposed in this paper follows an approach that Crommelin used to derive the distribution of the waiting time in an $M/D/N - \infty$ system [6]. His method has been adapted to the discrete–time equivalent $Geo/D/N - \infty$ system by Vicari and Tran–Gia [7].

This paper is structured as follows. Section 2 provides a description of the queueing system $M/D^b/N-\infty$. In Section 3, we propose a method of analysis for this system and derive performance measures of interest. We present some numerical results in Section 4 and summarize the paper in Section 5.

2 The $M/D^b/N - \infty$ queueing model

We consider the queueing system depicted in Figure 1. It consists of a queue of infinite length and N identical batch servers. Each of the servers has a batch capacity of b

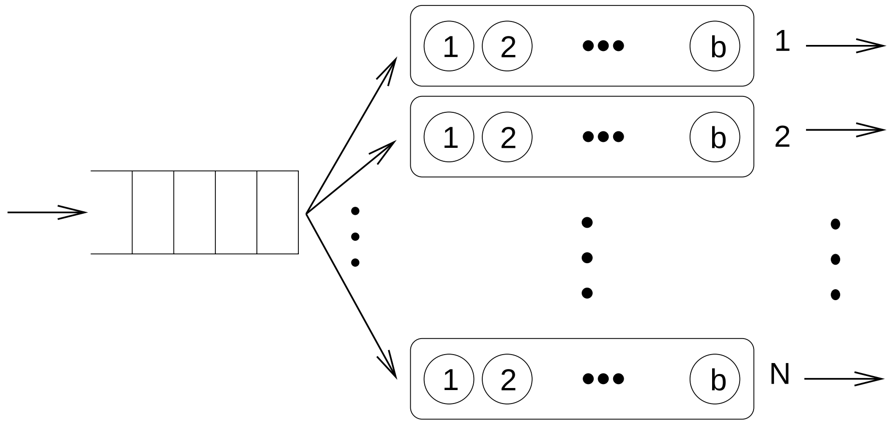

Figure 1: The queueing model

lots, i.e., a server can process b lots simultaneously. The servers are operated under a *full batch* strategy, i.e., a new service is not started until there is a full batch of size b available to process. Service time B is deterministic and identical for all N servers.

The interarrival times of lots arriving at the system follow an exponential distribution with mean $\mathrm{E}[A]$. Lots that find all servers busy and lots that have to wait for a full batch to form are queued in FIFO order. We define the system load as

$$\rho = \frac{B}{\mathrm{E}[A] \cdot N \cdot b}. \tag{1}$$

The queueing system is stable for a system load $\rho < 1$.

3 Performance analysis

In this section, we derive the distribution of the number of lots in queue at arbitrary time instants. The method we present follows an approach Crommelin [6] used to analyze the $M/D/N - \infty$ system.

Let the random variable Z describe the number of lots in the system at an arbitrary time instant t. Let Z' be the number of lots in the system at time $t + B$. Since at time $t + B$, all lots in service at time t have left the system, we get

$$Z' = \begin{cases} Z + \Gamma(B), & \text{if } Z < b \\ Z - b + \Gamma(B), & \text{if } b \leq Z < 2b \\ Z - 2b + \Gamma(B), & \text{if } 2b \leq Z < 3b \\ \quad\vdots & \quad\quad\vdots \\ Z - Nb + \Gamma(B), & \text{if } Nb \leq Z, \end{cases} \tag{2}$$

where $\Gamma(B)$ is the random variable denoting the number of arrivals during an interval of length B. $\Gamma(B)$ is Poisson distributed, hence $\gamma(B, k)$, the probability of having k arrivals during an interval of constant length B is

$$\gamma(B,k) \quad = \quad \frac{y^k}{k!}e^{-y}\,, \quad y = \frac{B}{\mathrm{E}[A]}\,, \quad k = 0,1,\ldots\,. \tag{3}$$

The distribution $z'(k)$ of Z' can be derived as follows:

$$z'(k) \quad = \quad \begin{cases} \sum_{i=0}^{N} z(i\cdot b+k) \circledast \gamma(B,k) & , \quad k = 0,\ldots,b-1 \\ z(N\cdot b+k) \circledast \gamma(B,k) & , \quad k \geq b\,. \end{cases} \tag{4}$$

Here, $z(k)$ denotes the distribution of Z and '$\circledast$' is the discrete convolution operation.

Setting $z(k) = z'(k)$ and applying Equation 4 iteratively, $z(k)$ will converge to the steady state distribution of the number of lots in system at arbitrary time instants.

Using $z(k)$, we can derive the mean number of lots in system as

$$\mathrm{E}[Z] \quad = \quad \sum_{k=0}^{\infty} k\cdot z(k)\,. \tag{5}$$

Since the mean number of lots in service is $N\cdot b\cdot\rho$, we get the mean number of lots in queue

$$\mathrm{E}[Q] \quad = \quad \mathrm{E}[Z] - N\cdot b\cdot\rho\,. \tag{6}$$

Applying *Little's Law* (cf. [8, p. 17]), we obtain the mean waiting time of lots before service as

$$\mathrm{E}[W] \quad = \quad \mathrm{E}[A]\cdot\mathrm{E}[Z]\,. \tag{7}$$

Waiting Time Distribution

We now derive the distribution of the waiting time of an arbitrary lot before receiving service. A waiting time of length t can be written as

$$t = mB + r\,, \quad r < B\,. \tag{8}$$

In order for the waiting time of a marked lot to be smaller than t, the following two conditions have to be fulfilled. First, t time units after the arrival of the marked lot, at most $Nb-1$ lots are allowed to be in queue, not including the marked lot. If we use Z_t to denote the number of lots in queue in front of the marked lot t time units after its arrival, we can write this condition as

$$Z_t \quad \leq \quad Nb-1\,. \tag{9}$$

Since we assume that all servers work continuously if there are lots to process, this condition can be rewritten as

$$Z_r \quad \leq \quad mNb + Nb - 1\,. \tag{10}$$

The second condition is that t time units after the arrival of the marked lot there are enough lots in queue to form a full batch. Denoting the cumulative arrival time of i lots by A_i, we can write this condition as

$$A_{b-\mathrm{mod}(Z,b)-1} \quad \leq \quad t\,. \tag{11}$$

Since the two conditions are independent of each other, the distribution function of the waiting time of a marked lot is

$$W(t) \quad = \quad Z_r(mNb+Nb-1)\cdot A_{b-\mathrm{mod}(Z,b)-1}(t)\,, \tag{12}$$

where $Z_r(k)$ is the distribution function of the number of lots in queue in front of the marked lot r time units after its arrival and $A_i(t)$ is the distribution function of the cumulative arrival time of i lots.

$Z_r(k)$ can be derived from the distribution of the number of lots in queue at arbitrary instants according to the following equation:

$$\begin{aligned} & Z_r(mNb+Nb-1) = \\ & \qquad Z(mNb+Nb-1) + \sum_{k=0}^{Nb-1} z(mNb+Nb+k)\cdot(1-S(\lfloor k/b\rfloor+1))\,, \end{aligned} \tag{13}$$

where $Z(k)$ is the distribution function of the number of lots in queue at arbitrary instants and $S(k)$ is the distribution function of the number S of servers that have completed service at time r. S is distributed according to a binomial distribution with parameters N and r/B, hence

$$S(k) \quad = \quad \sum_{i=0}^{k}\binom{N}{i}\left(\frac{r}{B}\right)^i\left(1-\frac{r}{B}\right)^{N-i}\,. \tag{14}$$

4 Numerical results

In this section, we present some numerical results for the $M/D^b/N-\infty$ queueing system. We compare six systems with an identical service time of $B=30$. The numbers of servers N and batch sizes b for the individual systems are given in Table 1.

Table 1: System parameters

System	1	2	3	4	5	6
N	1	2	3	4	6	12
b	12	6	4	3	2	1

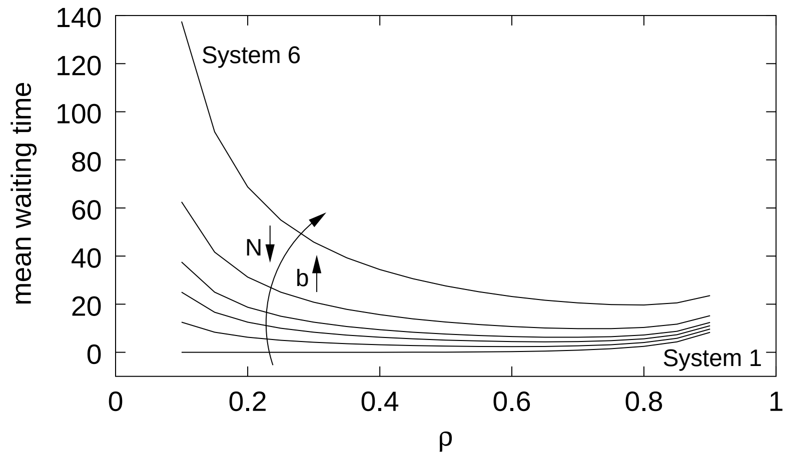

Figure 2: Mean waiting time

The mean waiting times for these systems are shown in Figure 2. Using the full batch strategy, large batch sizes lead to high waiting times, especially if ρ is small. This is due to the fact that lots have to wait longer for a full batch to form if the batch sizes are large and the arrival rate is low.

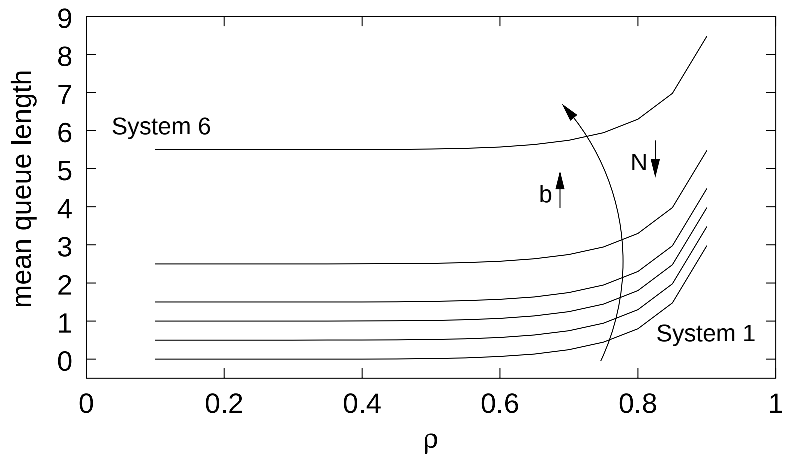

Figure 3: Mean queue length

The mean queue lengths, displayed in Figure 3, are almost constant for low values of ρ. However, the mean waiting times differ significantly when ρ is varied. Hence, estimating system performance just by considering the queue length can be somewhat misleading for this system.

In Figure 4, the complementary waiting time distribution functions for the six systems are given for a load of $\rho = 0.7$.

For larger number of servers and smaller batch capacities, the distributions exhibit a steeper descend. In summary, a system with a large number N of servers and smaller batch sizes b should be preferred to a system with few servers and larger batch size, if

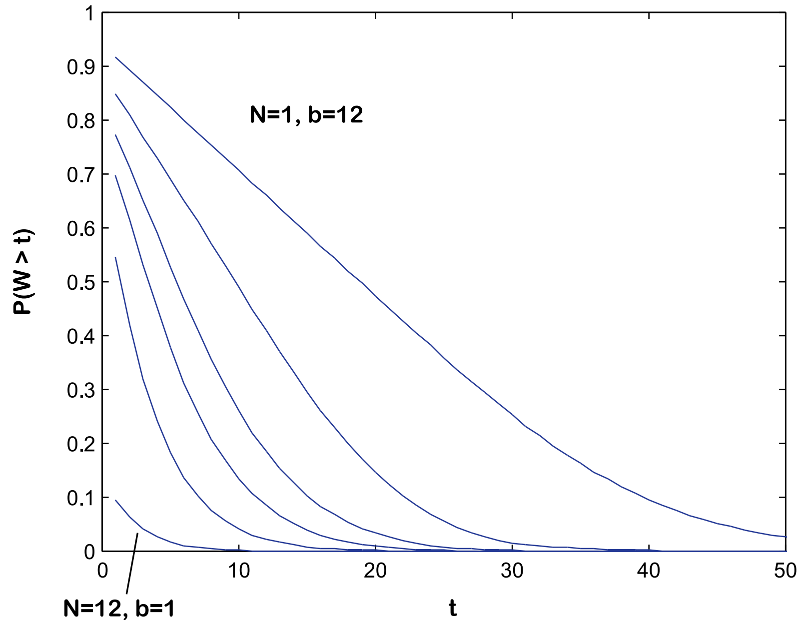

Figure 4: Complementary waiting time distribution function

small waiting times are desired.

5 Conclusion

We presented an iterative algorithm to analyze the $M/D^b/N-\infty$ queueing system. The algorithm is of low complexity, concerning both run time and memory usage. It allows to compute the distribution of the lots in queue and the distribution of the waiting time.

A number of extensions of the algorithm will have to be considered in the future. In addition to the full batch strategy, other strategies like the *greedy* rule or the *minimum batch size* rule have to be considered. The model may also be extended to include downtimes of the servers.

References

[1] P. Singer, "Furnaces evolving to meet diverse thermal processing needs," *Semiconductor International*, pp. 84–88, March 1997.

[2] M. F. Neuts, "A general class of bulk queues with poisson input," *Ann. Math. Stat.*, vol. 38, pp. 759–770, 1967.

[3] M. L. Chaudhry and J. G. C. Templeton, *A First Course in Bulk Queues*. New York: Wiley, 1983.

[4] H. Gold, *Performance Modelling of Batch Service Systems with Push and Pull manufacturing Management Policies.* PhD thesis, Universität Würzburg, Fakultät für Mathematik und Informatik, 1992.

[5] P. Tran-Gia, H. Gold, and H. Grob, "A batch service system operating in a pull production line," *Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 5/6, pp. 379–385, 1993.

[6] C. D. Crommelin, "Delay Probability Formulae," *Post Office Electrical Engeneer's Journal*, vol. 26, pp. 266–274, 1933.

[7] N. Vicari and P. Tran-Gia, "A numerical analysis of the $Geo/D/N$ queueing system," Forschungsbericht, Preprint-Reihe Nr. 151, Universität Würzburg, Institut für Informatik, Sept. 1996.

[8] L. Kleinrock, *Queueing Systems, Vol. 1: Theory.* New York: Wiley, 1975.