Exact Blocking Time Statistics for the Erlang Loss Model

Peter J. Smith, *Senior Member, IEEE*, Pawel A. Dmochowski, *Senior Member, IEEE*

Abstract**—The Erlang loss model is one of the fundamental tools in queueing theory with many applications to communications networks. For example, in a simple cellular voice network, the Erlang-B blocking formula is the traditional approach to model the proportion of time cellular base-stations are fully loaded in the busy hour. Such steady state results do not provide information on such important questions as: how likely is it that the blocking time in the busy hour exceeds some threshold. Hence we look in detail at the blocking time in the busy hour, or any finite period. We derive the exact distribution function and density as well as the moments and moment generating function of the blocking time, denoted X. In addition we derive the probabilities of zero blocking, $P(X = 0)$, and complete blocking, $P(X = 1$ hour$)$, in the busy hour.**

I. Introduction

The ELM is a basic model in queueing theory [1] and is of fundamental importance in the analysis of blocking in telephone traffic [2]. The ELM and the associated blocking probability (Erlang-B) are discussed in virtually all texts on networks and in the past, Erlang-B formulae have been widely used in dimensioning networks, such as cellular voice networks. Even with the advent of multi-media networks and mixtures of data and voice, such results remain useful guides and are widely used due to their simplicity in both industry [3] and in research [4]. Erlang-B results are still finding applications today [5] and the ELM remains the basis of many traffic models [6]. The ELM is also commonly used as a basis for extension to more complex situations. For example, multi-dimensional queues built on the ELM which handle different traffic types have been applied to cellular OFDMA [7] and cognitive radio scenarios [8]. Although the ELM is a general queueing model with many applications, for ease of exposition we discuss the ELM in the concrete context of a cellular radio network. Hence, the channels available at the base station are the servers and the cellular calls are the customers.

The dominant use of the ELM in virtually all applications is the simple closed form blocking probability, the Erlang-B result, that stems from the ELM. However, such steady state results do not give information on how likely it is to observe long periods of blocking. Hence, our aim is a complete characterization of the blocking time, X, in a fixed interval, assuming the ELM holds. Such results are not available, although other transient analyses of the ELM have appeared [1], [2], [9]–[11]. Hence, we derive exact expressions for the distribution function, density and moments of the blocking time. In addition, we compute the moment generating function (mgf) of the blocking time and the probabilities of zero blocking and complete blocking for the hour. A sketch proof of the distribution function appeared previously in Tunnicliffe et. al. [12]. In this paper, we extend this work and give a complete derivation of the distribution function and a numerically stable approach to computation as well as the completely new results on blocking probabilities, moments, the density and the mgf. The method is based on an example of the uniformization technique given in Ross [13, p.174].

P. J. Smith is with the Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (email: peter.smith@canterbury.ac.nz).

P. A. Dmochowski is with the School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand (email: pdmochowski@ieee.org).

II. Blocking Time Distribution Function

Consider the ELM with M channels, Poisson call arrivals (rate λ) and exponential call times (rate μ). The model assumes an infinite population of potential callers and blocked calls are lost. Let $X(\tau)$ be the continuous time Markov chain given by the number of channels in use at time τ for $\tau \in [0, t]$. Also, define $S_i(t)$ to be the total busy time (time in state M) in $[0, t]$, given the initial state $X(0) = i$. The unconditional busy time is denoted $S(t)$ with the distribution function

$$\begin{aligned} F(x) &= P(S(t) \leq x) \\ &= \sum_{i=0}^{M} P(S_i(t) \leq x) P(X(0) = i). \end{aligned} \tag{1}$$

We require an expression for the distribution function in (1) above. If we assume the system is in steady state at $\tau = 0$, then the initial state probability is given by the Erlang-B or truncated Poisson formula

$$P(X(0) = i) = q_i(\lambda/\mu) \Big/ \sum_{j=0}^{M} q_j(\lambda/\mu), \quad i = 0, 1, \dots, M, \tag{2}$$

where $q_i(\lambda) = \lambda^i \exp(-\lambda)/i!$. The key to computing the first probability in (1) is to use a technique known as uniformization [13, p.174] which makes all states have the same rate of transition by introducing intra-state transitions. This section applies this methodology to the ELM.

The uniformized process is completely equivalent to the old one but has different transition probabilities [13, p.175],

$$p^*_{ij} = \begin{cases} 1 - v_i/v & j = i \\ (v_i/v) p_{ij} & j \neq i, \end{cases} \tag{3}$$

where p_{ij} are the transition probabilities of the original process, v_i is the rate the original process leaves state i and $v = \max(v_0, v_1, \dots, v_M)$ is the common rate at which the new process leaves any state. For a birth and death process, like the ELM, these terms are well known [13, p.175,176],

$$v_i = \begin{cases} \lambda + i\mu & i = 0, 1, \ldots, M-1 \\ M\mu & i = M, \end{cases} \tag{4}$$

$$v = \lambda + (M-1)\mu, \text{ assuming } \lambda > \mu, \tag{5}$$

$$\left.\begin{aligned} p_{i,i+1} &= \lambda/(\lambda + i\mu) = 1 - p_{i,i-1}, \quad i = 1, 2, \ldots, M-1, \\ p_{0,1} &= p_{M,M-1} = 1. \end{aligned}\right\} \tag{6}$$

Using the uniformized process, we can compute the first probability in (1) by conditioning on $N(t)$, the number of transitions in $[0, t]$ made by the new process, and noting that $N(t)$ is a Poisson process with rate v [13, p.175]. Hence,

$$P(S_i(t) \le x) = \sum_{n=0}^{\infty} P(S_i(t) \le x | N(t) = n) q_n(vt).$$

Now, defining $Z(t)$ to be the total number of visits to the busy state in $[0, t]$ and $X_n(k)$ to be the total time spent in the busy state during k visits out of n transitions we have

$$P(S_i(t) \le x) = \sum_{n=0}^{\infty} q_n(vt) \tag{7}$$
$$\times \sum_{k=0}^{n+1} P(Z(t) = k | X(0) = i, N(t) = n) P(X_n(k) \le x).$$

From [13, p.178], the last probability in (7) is the binomial tail probability

$$P(X_n(k) \le x) = \begin{cases} \sum_{i=k}^{n} \binom{n}{i} \left(\frac{x}{t}\right)^i \left(1 - \frac{x}{t}\right)^{n-i}, & k \le n, x < t \\ 0, & k = n+1, x < t \\ 1, & x = t. \end{cases} \tag{8}$$

The conditional probability for $Z(t)$ in (7) is harder and closed form expressions seem infeasible for anything other than the smallest cases, $M = 2$ or 3 say. Hence, we use a first step analysis [14, p.84] to create a recursion. Denoting $q(k, i, n) = P(Z(t) = k | X(0) = i, N(t) = n)$ for compactness this gives, for $k = 0, 1, \ldots, n+1$, and $n = 0, 1, \ldots$

$$\left.\begin{aligned} q(k, i, n) &= \sum_{j=i-1}^{i+1} p^*_{ij} q(k, j, n-1), \quad i = 1, \ldots, M-1 \\ q(k, 0, n) &= \sum_{j=0}^{1} p^*_{0j} q(k, j, n-1), \\ q(k, M, n) &= \sum_{j=M-1}^{M} p^*_{Mj} q(k-1, j, n-1), \end{aligned}\right\} \tag{9}$$

which can be used as a recursion to the boundary values

$$\left.\begin{aligned} q(0, i, 0) &= 1, \quad i = 0, 1, \ldots, M-1, \\ q(0, M, n) &= 0, \quad n = 0, 1, \ldots, \\ q(n+1, M, n) &= (p^*_{MM})^n, \quad n = 0, 1, \ldots, \\ q(n+1, i, n) &= 0, \quad n = 0, 1, \ldots, \quad i = 0, 1, \ldots, M-1. \end{aligned}\right\} \tag{10}$$

As an example, consider the first equation in (9). Starting in state i, the process can only move to states $j = i-1$, $j = i$ or $j = i+1$. Hence, there are 3 possible transitions from i to j with probability p^*_{ij}. Conditional on the actual transition made, from state j there must still be k visits to the busy state in the reduced number of $n-1$ remaining transitions. This has probability $q(k, j, n-1)$ and the first equation in (9) follows from this first step analysis.

Hence, the distribution of the busy time is given exactly by equations (1)–(10) and is summarised below:

$$P(S(t) \le x) = \left\{ \sum_{i=0}^{M} q_i(\lambda/\mu) \sum_{n=0}^{\infty} q_n(vt) \sum_{k=0}^{n+1} q(k, i, n) P(X_n(k) \le x) \right\} \times \left[\sum_{j=0}^{M} q_j(\lambda/\mu) \right]^{-1}. \tag{11}$$

The structure of (11) is simple for numerically stable summation since all terms are probabilities and hence, the summation is over positive bounded terms. The only complexity in (11) is the computation of the $q(k, i, n)$ terms. However, the 3-D array, $q(k, i, n)$, can be efficiently computed using a layered approach afforded by the recursion relationships in (9). The lower boundary in the k, i, n space is specified by (10). Starting with $n = 0$, the array can be constructed by populating each k, i plane (limited by $k \le n+1$ and $i \le M$) using (9). Each step in (9) simply involves a summation of 2 or 3 neighboring elements, weighted by the transition probabilities in (3). After a k, i plane is complete, the recursion proceeds upwards along the n axis and another k, i plane is populated.

III. Probabilities, Density and Moments

The distribution is of mixed type with a continuous density over $(0, t)$ and non-zero probabilities at zero and t, given by

$$P(S(t) = 0) = \left\{ \sum_{i=0}^{M-1} q_i(\lambda/\mu) \sum_{n=0}^{\infty} q_n(vt) q(0, i, n) \right\} \times \left[\sum_{j=0}^{M} q_j(\lambda/\mu) \right]^{-1}, \tag{12}$$

and

$$P(S(t) = t) = \left\{ q_M(\lambda/\mu) e^{-M\mu t} \right\} \Big/ \sum_{j=0}^{M} q_j(\lambda/\mu) \tag{13}$$

Equation (12) follows from (11) with $x = 0$. Since no time is spent in the busy state, the initial state cannot be M and so the initial summation in (12) is from 0 to $M-1$ rather than from 0 to M. Also, there cannot be any visits to state M so that only $q(0, i, n)$ is relevant. Finally, given that there are no busy state visits, $P(X_n \le 0) = 1$ and (12) is seen as a special case of (11). Obtaining (13) is easier as the system must start in the busy state which occurs with probability $q_M(\lambda/\mu)/\sum_{j=0}^{M} q_j(\lambda/\mu)$. After this, no departures are allowed for a period of time t. Since the probability of an individual departure exceeding time t is $e^{-\mu t}$ the probability that no calls leave before time t is $e^{-M\mu t}$ and (13) follows.

Note that the only term in (11) involving x is $P(X_n(k) \leq x)$. Hence, when computing the density and moments it suffices to compute $g'(x)$ and $\int_0^t x^r g'(x)\ dx$ respectively, where $g(x) = \left(\frac{x}{t}\right)^i \left(1-\frac{x}{t}\right)^{n-i}$. Using the results

$$g'(x) = \left(\frac{x}{t}\right)^{i-1}\left(1-\frac{x}{t}\right)^{n-i-1}\left(\frac{1}{t}\right)\left\{i - \frac{nx}{t}\right\},$$

$$\int_0^t x^r g'(x)\ dx = \begin{cases} -rt^r B(r+i, n-i+1) & i<n \\ \frac{nt^r}{n+r} & i=n, \end{cases}$$

we can express the continuous part of the density as

$$\begin{aligned} f(x) = \Bigg\{ &\sum_{i=0}^{M} q_i(\lambda/\mu) \sum_{n=0}^{\infty} q_n(vt) \sum_{k=0}^{n} q(k,i,n) \\ &\times \sum_{j=k}^{n} \binom{n}{j}\left(\frac{x}{t}\right)^{j-1}\left(1-\frac{x}{t}\right)^{n-j-1}\left(\frac{1}{t}\right)\left(j-\frac{nx}{t}\right)\Bigg\} \\ &\times \left[\sum_{j=0}^{M} q_j(\lambda/\mu)\right]^{-1} \end{aligned} \tag{14}$$

for $0 < x < t$. The moments are given by

$$\begin{aligned} E[S^r(t)] = \Bigg\{ &\sum_{i=0}^{M} q_i(\lambda/\mu) \sum_{n=0}^{\infty} q_n(vt) \sum_{k=0}^{n} q(k,i,n) \\ &\times \left[\sum_{j=k}^{n-1}\binom{n}{j}(-rt^r B(r+j, n-j+1)) + nt^r/(n+r)\right]\Bigg\} \\ &\times \left[\sum_{j=0}^{M} q_j(\lambda/\mu)\right]^{-1} + t^r P(S(t)=t). \end{aligned} \tag{15}$$

The mgf of $S(t)$ can also be derived from

$$\begin{aligned} M_{S(t)}(\omega) &= E\left[e^{\omega S(t)}\right] \qquad (16) \\ &= \int_0^t f(x) e^{\omega x}\ dx + P\left(S(t)=0\right) + e^{\omega t} P\left(S(t)=t\right). \end{aligned}$$

The probabilities in (16) are given by (12) and (13). Defining $G_{j,n}(x) = \left(\frac{x}{t}\right)^{j-1}\left(1-\frac{x}{t}\right)^{n-j-1}\left(j-\frac{nx}{t}\right)$, the integral in (16) can be written as

$$\begin{aligned} &\sum_{i=0}^{M} q_i(\lambda/\mu) \sum_{n=0}^{\infty} q_n(vt) \sum_{k=0}^{n} q(k,i,n) \\ &\times \sum_{j=k}^{n}\binom{n}{j}\left(\frac{1}{t}\right) H_{j,n}(\omega)\left[\sum_{j=0}^{n} q_j(\lambda/\mu)\right]^{-1}, \end{aligned}$$

where

$$H_{j,n}(\omega) = \int_0^t G_{j,n}(x) e^{\omega x}\ dx.$$

The terms, $H_{j,n}(\omega)$, are straightforward to compute by expanding the binomial term in $G_{j,n}(x)$ and using standard integrals in [15, p.364]. This approach gives the result

$$H_{j,n}(\omega) = \begin{cases} 0 & j=n=0 \\ nt(-\omega t)^{-n}\gamma(n, -\omega t) & j=n>0 \\ -nte^{\omega t}(\omega t)^{-n}\gamma(n, \omega t) & j=0, n>0 \\ t\sum_{r=0}^{n-j-1}\binom{n-j-1}{r}(-1)^r \times & \\ \left\{j(-\omega t)^{-(r+j)}\gamma(r+j, -\omega t)\right. & \\ \left.-n(-\omega t)^{-(r+j+1)}\gamma(r+j+1, -\omega t)\right\} & \text{otherwise} \end{cases}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function.

Hence, we have a complete characterisation of the busy time in $[0,t]$ via the distribution function (11), discrete probabilities (12–13), density (14), moments (15) and the mgf (16).

IV. SIMULATION RESULTS

We present Matlab simulation results in order to validate the analytical expressions derived in Sections II and III. Monte Carlo simulations were used to evaluate the empirical blocking time CDF and mean blocking time. The system was modelled using the standard exponential inter-arrival times and departure times for the ELM. This created a single realisation of the ELM for one finite time period, which was replicated 10^4 times to provide simulation results. These were compared with the theoretical results in (11) and (15), respectively.

Fig. 1 shows the CDF of the blocking time for the scenarios: $M = 3$, $\lambda = 1.5$, $\mu = 0.3$ and $M = 3$, $\lambda = 5$, $\mu = 1$. While these parameter choices are somewhat unrealistic, they are chosen to demonstrate the agreement of the analytical results for both the continuous part of the CDF and the zero and full blocking endpoint probabilities given by (12) and (13). We note good agreement between simulations and analysis.

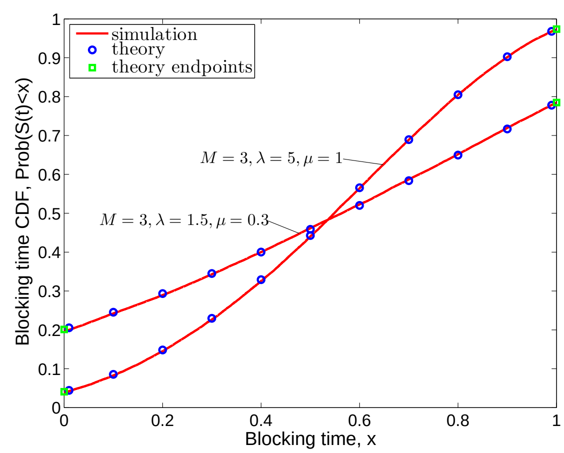

Fig. 1. Blocking time CDFs for $t = 1$.

The effect of the number of channels on the blocking time CDF is shown in Fig. 2, which shows results for an arrival rate of $\lambda = 150$ and a call service rate of $\mu = 10$. Noting the logarithmic scale on the abscissa, we observe a rapid reduction in the blocking probability CDF with increasing M. For example, a system with $M = 25$ and $M = 20$ channels results in a five-fold increase in the 90th percentile from 2% to 10% blocking. The analytical results obtained via (11) are in close agreement with the simulations.

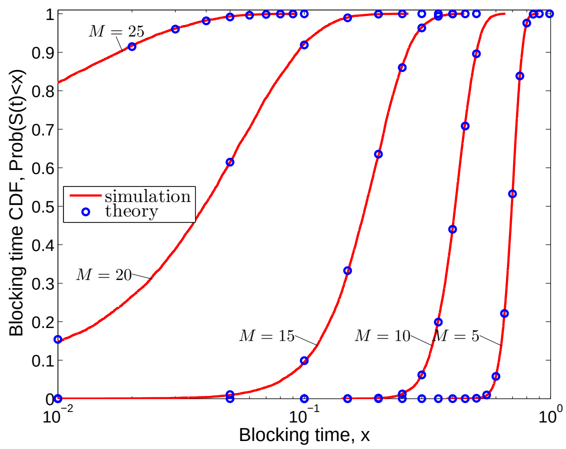

Fig. 2. Blocking time CDFs for varying numbers of channels ($t = 1$, $\lambda = 150$, $\mu = 10$).

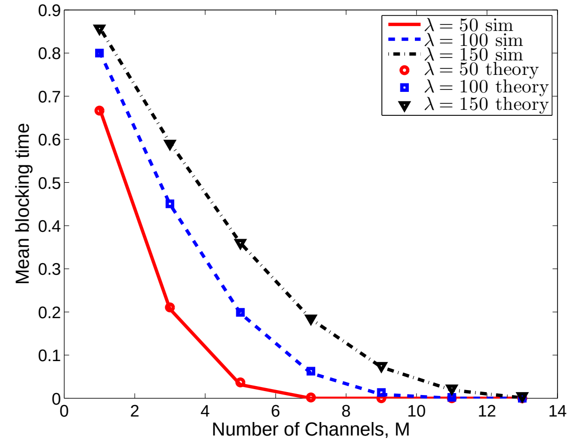

Fig. 4. Mean blocking time versus M ($t = 1$, $\mu = 25$).

Fig. 3 shows the complementary CDF (CCDF) of the blocking time for varying service rates, μ, with $\lambda = 150$ and $M = 10$. We use a log-scale to highlight the upper tail of the blocking time distribution. For $\mu = 25$, the Erlang-B blocking probability is 4% so the traditional approach to quantifying the system $\mu = 25$, $\lambda = 150$, $M = 10$ is via this steady state blocking probability of 4%. For this system there is a 10% chance that more than 8% of the busy hour is blocked and that there is a 3% chance that more than 10% of the hour is blocked. Hence, it is not unusual for periods of blocking to occur which are 2-3 times greater than the mean. The CCDF results allow us to quantify this information in a precise way.

Finally, Fig. 4 shows the mean blocking time for $\mu = 25$, where the theoretical values obtained using (15) are verified via simulation. Note that, as expected, the mean blocking times from (15) also agree exactly with the Erlang-B results. However, (15) is more general and also provides the higher order moments and the variance in particular.

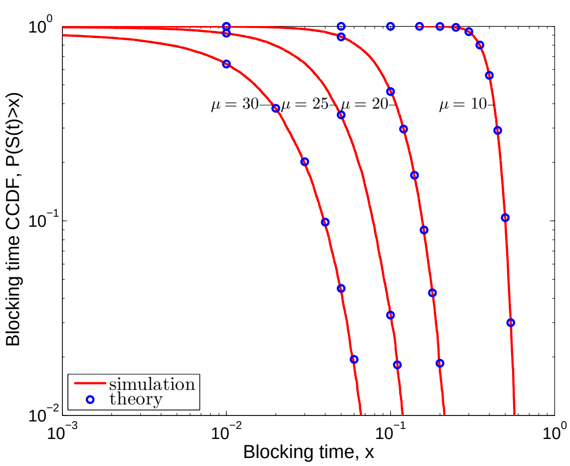

Fig. 3. Blocking time CCDFs for different call times ($t = 1$, $\lambda = 150$, $M = 10$).

V. Conclusions

We have provided a complete study of the finite time blocking characteristics of the ELM. In particular, we have investigated the distribution of the blocking time in a fixed period, $[0, t]$, assuming the ELM holds. Exact expressions are given for the distribution function, density, moments and mgf of the blocking time. The probabilities of zero and complete blocking are also given. To the best of the authors' knowledge this is the first time such expressions have been derived. Simulation results showing close agreement with the theoretical results were presented for a range of system parameters.

References

[1] L. Takács, *Introduction to the Theory of Queues*. New York: Oxford University Press, 1962.

[2] V. Beneš, *Mathematical Theory of Connecting Networks and Telephone Traffic*. New York: Academic Press, 1965.

[3] "Capacity management and optimization of voice traffic," Cisco, Tech. Rep., Nov. 2007.

[4] T. Misuth and I. Baronak, "Application of Erlang B model in modern VoIP networks," in *Telecommunications and Signal Processing (TSP), 2011 34th International Conference on*, Aug. 2011, pp. 235 –239.

[5] B. Zhang, R. Iyer, and K. Kiasaleh, "Reverse link Erlang capacity of OFDMA wireless systems with adaptive resource allocation," in *IEEE Wireless Com. and Net. Conf. (WCNC)*, April 2006, pp. 2106–2109.

[6] S. Batabyal and S. Das, "Distance dependent call blocking probability, and area Erlang efficiency of cellular networks," in *IEEE 75th Vehicular Technology Conference (VTC Spring)*, May 2012, pp. 1–5.

[7] G. Joshi, H. Maral, and A. Karandikar, "Downlink Erlang capacity of cellular OFDMA," in *National Conf. on Communications (NCC)*, Jan. 2011, pp. 1–5.

[8] A. Firag, P. Smith, P. Dmochowski, and M. Shafi, "Analysis of the M/M/N/N queue with two types of arrival process: Applications to future mobile radio systems," *Journal of Applied Mathematics*, vol. 2012, no. Article ID 123808, pp. 1–14, 2012.

[9] C. Knessl, "On the transient behaviour of the M/M/m/m loss model," *Communications in Statistics—Stochastic Models*, vol. 6, no. 4, pp. 749–776, 1990.

[10] D. Mitra and A. Weiss, "The transient behaviour in Erlang's model for large trunk groups and various traffic conditions," in *Proc. the 12th ITC*, Torino, Italy, June 1998, pp. 1367–1374.

[11] P. Gazdzicki, I. Lambadaris, and R. Mazumdar, "Blocking probabilities for large multirate Erlang loss systems," *Adv. Appl. Prob.*, vol. 25, pp. 997–1009, 1993.

[12] P. Smith, A. Sathyendran, and A. Murch, "Analysis of traffic distribution in cellular networks," in *IEEE 49th Vehicular Technology Conf. (VTC)*, vol. 3, July 1999, pp. 2075–2079.

[13] S. Ross, *Stochastic Processes*. New York: John Wiley and Sons, Inc., 1996.

[14] M. Pinsky and S. Karlin, *An Introduction to Stochastic Modeling*, 4th ed. Academic Press, 2011.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th ed. San Diego, CA: Academic Press, 2007.