



**HAL**  
open science

# The Erlang Model with non-Poisson Call Arrivals

Thomas Bonald

► **To cite this version:**

Thomas Bonald. The Erlang Model with non-Poisson Call Arrivals. Sigmetrics / Performance 2006 - Joint International Conference on Measurement and Modeling of Computer Systems, Jun 2006, Saint Malo, France. hal-01274120

**HAL Id: hal-01274120**

**<https://hal.science/hal-01274120>**

Submitted on 17 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Erlang Model with non-Poisson Call Arrivals

Thomas Bonald\*  
France Telecom R&D  
Issy-les-Moulineaux, France

## Abstract

The Erlang formula is known to be insensitive to the holding time distribution beyond the mean. While calls are generally assumed to arrive as a Poisson process, we prove that it is in fact sufficient that users generate *sessions* according to a Poisson process, each session being composed of a random, finite number of calls and idle periods. A key role is played by the retrial behavior in case of call blocking. We illustrate the results by a number of examples.

## 1 Introduction

Since its publication in 1917, the Erlang formula has proved very useful for dimensioning telephone networks [8]. It determines the required number of telephone lines given a prediction of expected demand and a target blocking probability. A key property of the Erlang formula is its insensitivity: the blocking probability does not depend on the holding time distribution beyond the mean [18]. Traffic is in fact characterized by a unique parameter, the *traffic intensity*, which is defined as the product of the call arrival rate and the mean holding time. This makes the Erlang formula both simple to apply and robust to changes in fine traffic characteristics, and explains its enduring success.

The only assumption required by the Erlang model is that calls arrive as a Poisson process. This property naturally arises when calls are generated independently by a large number of users. Basically, if each user generates at most one call in a period of time where the system is in steady state (e.g., the busy hour), the arrival time of each call is uniformly distributed over that period so that call arrivals indeed form a Poisson process. It is frequent, however, that users generate a sequence of calls instead of a single, isolated call. The call arrival process is then *not* a Poisson process. Defining a session as the sequence of calls generated by the same user, it only holds that the *session* arrival process is a Poisson process. We prove in this paper that the Erlang formula still holds in this case, under some mild assumptions on the retrial behavior in case of call blocking.

Specifically, we consider two types of retrial behavior in case of call blocking: either the session goes on as if the call were accepted and terminated instantaneously, or the call is reattempted with a fixed probability  $r$  after an idle period of random duration, the session going on with probability  $1 - r$ . In the latter case, the duration of the idle period between two attempts has the same distribution as the idle period between the end of the previous call and the first attempt of the considered call. If the considered call is the first of a session, there is no retrial and the session is lost with probability  $r$  and goes on with probability  $1 - r$ .

Under these assumptions, the Erlang formula is insensitive to *all* traffic characteristics beyond the traffic intensity. The number of calls per session, the call durations and the idle durations may have arbitrary distributions. There may be arbitrary correlation between these random variables. The blocking probability of a call is given by the Erlang formula, independently of the characteristics of the session it belongs to and of its location in this session (e.g., the first, second or last call of the session). The only requirement is that sessions are mutually independent and arrive as a Poisson process.

---

\*Also affiliated with École Normale Supérieure, France.

Sessions may not arrive as a Poisson process, however. This is the case when the number of users is relatively low so that the session arrival intensity is not independent of the number of ongoing sessions (the higher the number of ongoing sessions, the less likely the arrival of new sessions). All sessions must then be considered as permanent, the user activity being determined by the ratio of mean call duration to mean idle duration. This is the well-known Engset model [6, 2]. For equal traffic intensities, the Engset formula gives a lower blocking rate than the Erlang formula and tends to the Erlang formula when the number of users tends to infinity.

The insensitivity property of the Erlang formula is satisfied by the Engset formula under the same assumptions on the retrial behavior. In particular, while call durations and idle durations are generally assumed to be independent and identically distributed, there may in fact be arbitrary correlation between these random variables. Note that the two types of retrial behavior described above are equivalent for the Engset model under i.i.d. assumptions but differ in the presence of correlation.

The combination of the Erlang model and the Engset model is sufficiently general to represent any traffic conditions. In both cases, the blocking probability depends on the traffic intensity only, which makes these models even more robust than generally believed. The insensitivity property extends to general loss networks [12, 16]. Such models naturally arise in the performance analysis of wired and wireless networks when physical or virtual constant bit rate circuits are used to transmit various traffic streams like data, voice, audio and video flows. While such a traffic mix may result in a bursty flow arrival process, this does not affect the blocking probability provided flows are generated within independent sessions.

We start with preliminary results on queueing networks that are used in Section 3 to prove the insensitivity of the Erlang formula. The two types of retrial behavior mentioned above, referred to as jump-over blocking and random retrials, are presented in Section 4. Section 5 is devoted to some extensions of the main result, including the insensitivity of the Engset formula. The results are illustrated by a number of examples in Section 6. Section 7 concludes the paper.

## 2 Preliminary results

### 2.1 A queueing network

Consider a network of  $N$  infinite server queues. Customers arrive at queue  $i$  as a Poisson process of intensity  $\nu_i$ , with

$$\nu \stackrel{\text{def}}{=} \sum_{i=1}^N \nu_i > 0.$$

Service times at queue  $i$  are independent, exponentially distributed of mean  $1/\mu_i$ , for some  $\mu_i > 0$ . Once served at queue  $i$ , customers move to queue  $j$  with probability  $p_{ij}$  and leave the network with probability:

$$p_i \stackrel{\text{def}}{=} 1 - \sum_{j=1}^N p_{ij}.$$

All customers eventually leave the network. The arrival rate  $\lambda_i$  at queue  $i$  including arrivals from other queues is then uniquely defined by the traffic equations:

$$\lambda_i = \nu_i + \sum_{j=1}^N \lambda_j p_{ji}. \tag{1}$$

Summing these equations, we get the traffic conservation equation:

$$\nu = \sum_{i=1}^N \lambda_i p_i. \tag{2}$$

The traffic intensity at queue  $i$  is defined by  $\rho_i = \lambda_i/\mu_i$ .

Let  $X(t)$  be the  $N$ -dimensional vector whose  $i$ -th component  $X_i(t)$  is the number of customers in queue  $i$  at time  $t$ . It is well-known that the stationary distribution of the Markov process  $X(t)$  has a product form. We give a proof of this result stated as Theorem 1 to show the similarity with the proof of Theorem 2 below. In the rest of the paper, we denote by  $e_i$  the  $N$ -dimensional vector whose  $i$ -th component is equal to 1 and other components are equal to 0.

**Theorem 1** *The stationary distribution of the Markov process  $X(t)$  is given by:*

$$\pi(x) = \pi(0) \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!}$$

with

$$\pi(0) = \prod_{i=1}^N e^{-\rho_i}.$$

*Proof.* It follows from the traffic equations (1) and (2) that for all queues  $i$  and all states  $x \in \mathbb{N}^N$ :

$$\pi(x + e_i)\mu_i(x_i + 1) = \pi(x)\nu_i + \sum_{j=1}^N \pi(x + e_j)\mu_j(x_j + 1)p_{ji}$$

and

$$\pi(x) \sum_{i=1}^N \nu_i = \sum_{i=1}^N \pi(x + e_i)\mu_i(x_i + 1)p_i.$$

The equilibrium equations follow by summation:

$$\begin{aligned} \pi(x) \sum_{i=1}^N (\nu_i + \mu_i x_i) &= \\ \pi(x - e_i)\nu_i + \sum_{i,j=1}^N \pi(x + e_j - e_i)\mu_j(x_j + 1)p_{ji}, \end{aligned}$$

where we use the convention that  $\pi(x) = 0$  if  $x \notin \mathbb{N}^N$ . □

## 2.2 Restriction of the state space

Consider now a set of admissible states,  $\mathcal{A} \subset \mathbb{N}^N$ . We assume that  $\mathcal{A}$  is coordinate convex in the sense that if  $x \in \mathcal{A}$  then  $y \in \mathcal{A}$  for all vectors  $y$  such that  $y \leq x$  component-wise. Customers arrive at queue  $i$  as a Poisson process of intensity  $\nu'_i(x)$  in state  $x$ , with  $\nu'_i(x) = 0$  if  $x + e_i \notin \mathcal{A}$ . Once served at queue  $i$ , customers move to queue  $j$  with probability  $p'_{ij}(x)$  in state  $x$ , with  $p'_{ij}(x) = 0$  if  $x - e_i + e_j \notin \mathcal{A}$ , and leave the network with probability  $p'_i(x)$ , with

$$p'_i(x) = 1 - \sum_{j=1}^N p'_{ij}(x).$$

Let  $\lambda'_i(x)$  be the arrival rate of customers to queue  $i$  in state  $x$ . The corresponding traffic equations are:

$$\lambda'_i(x) = \nu'_i(x) + \sum_{j=1}^N \lambda'_j(x)p'_{ji}(x + e_j). \quad (3)$$

Summing these equations, we get the corresponding traffic conservation equation:

$$\sum_{i=1}^N \nu'_i(x) = \sum_{i=1}^N \lambda'_i(x) p'_i(x + e_i). \quad (4)$$

Note that these equations may not have a unique solution. If  $\nu'_i(x) = 0$  for all  $i$  in some state  $x$  for instance, the arrival rates are defined up to a multiplicative constant (the network is then locally a closed queueing network).

Let  $X'(t)$  be the  $N$ -dimensional vector whose  $i$ -th component  $X'_i(t)$  is the number of customers in queue  $i$  at time  $t$ . We have the following key result:

**Theorem 2** *If the arrival rates defined by:*

$$\lambda'_i(x) = \lambda_i \text{ if } x + e_i \in \mathcal{A}, \quad \lambda'_i(x) = 0 \text{ otherwise}$$

*are a solution of the traffic equations (3) for all queues  $i$  and all states  $x \in \mathbb{N}^N$ , then the stationary distribution  $\pi'$  of the Markov process  $X'(t)$  is the restriction of  $\pi$  to the set of admissible states  $\mathcal{A}$ , that is:*

$$\pi'(x) = \pi'(0) \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \text{ if } x \in \mathcal{A}, \quad \pi'(x) = 0 \text{ otherwise,}$$

with

$$\pi'(0) = \left( \sum_{x \in \mathcal{A}} \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \right)^{-1}.$$

*Proof.* It follows from the traffic equations (3) and (4) that for all queues  $i$  and all states  $x \in \mathbb{N}^N$ ,

$$\begin{aligned} \pi'(x + e_i) \mu_i(x_i + 1) &= \pi'(x) \nu'_i(x) \\ &+ \sum_{j=1}^N \pi'(x + e_j) \mu_j(x_j + 1) p'_{ji}(x + e_j) \end{aligned}$$

and

$$\pi'(x) \sum_{i=1}^N \nu'_i(x) = \sum_{i=1}^N \pi'(x + e_i) \mu_i(x_i + 1) p'_i(x + e_i).$$

The equilibrium equations follow by summation:

$$\begin{aligned} \pi'(x) \sum_{i=1}^N (\nu'_i(x) + \mu_i x_i) &= \pi'(x - e_i) \nu'_i(x - e_i) \\ &+ \sum_{i,j=1}^N \pi'(x + e_j - e_i) \mu_j(x_j + 1) p'_{ji}(x + e_j - e_i), \end{aligned}$$

where we use the convention that  $\pi(x) = 0$  if  $x \notin \mathbb{N}^N$ . □

The following ‘Departures See Time Average’ property is a direct consequence of Theorem 2. A queue  $i$  is said to be *free* if  $x + e_i \in \mathcal{A}$  for all admissible states  $x \in \mathcal{A}$ .

**Corollary 1** *Under the conditions of Theorem 2, customers leaving any free queue (either moving to another queue or leaving the network) see the network in steady state immediately after their departure.*

*Proof.* The probability  $\pi'_i(x)$  that customers leaving free queue  $i$  see the network in state  $x$  immediately after their departure is proportional to  $\pi'(x + e_i)\mu_i(x_i + 1)$ , so that:

$$\pi'_i(x) = \frac{\pi'(x + e_i)\mu_i(x_i + 1)}{\sum_{y \in \mathcal{A}} \pi'(y + e_i)\mu_i(y_i + 1)}.$$

Now it follows from Theorem 2 that for all  $x \in \mathcal{A}$ ,

$$\pi'(x + e_i)\mu_i(x_i + 1) = \pi'(x)\lambda_i.$$

We deduce:

$$\pi'_i(x) = \frac{\pi'(x)\lambda_i}{\sum_{y \in \mathcal{A}} \pi'(y)\lambda_i} = \pi'(x).$$

□

### 3 Insensitivity of the Erlang formula

The above results are now used to prove the insensitivity of the Erlang formula. We first show that the queueing network introduced in §2.1 can represent any traffic characteristics, with arbitrary distributions of the number of calls per session, call durations and idle durations. There may be arbitrary correlation between these random variables. The only required assumption is that sessions are independent and arrive as a Poisson process.

#### 3.1 Traffic characteristics

We decompose each call and each idle period in a finite, random number of exponential phases, as illustrated by Figure 1. The corresponding distributions, generally referred to as phase-type distributions or Cox distributions, are known to form a dense subset within the set of all distributions with real, non-negative support.

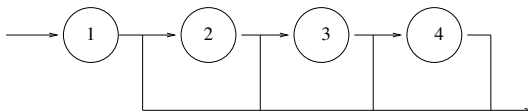


Figure 1: A Cox distribution with four exponential phases: after the completion of phase  $i$ ,  $i = 1, 2, 3$ , the service enters phase  $i + 1$  with probability  $q_i$  and ends with probability  $1 - q_i$ .

Each infinite-server queue of the network described in §2.1 represents the phase of a call or an idle period. We denote by  $\mathcal{C}$  the set of queues corresponding to *call* phases, by  $\mathcal{I}$  the set of queues corresponding to *idle* phases. These sets form a partition of the set of queues  $\{1, \dots, N\}$ . By convention, all sessions start and end with a call (and not an idle period). We denote by  $\mathcal{S} \subset \mathcal{C}$  the set of call phases corresponding to the starting phase of a session. These are the only queues  $i$  for which  $\nu_i > 0$ .

We assume that queues are numbered in such a way that phases of successive calls and idle periods of the same session correspond to increasing indices. Specifically, for any call phase  $i \in \mathcal{C}$ , we let  $p_{ij} = 0$  for all  $j$  except if  $j$  is the following phase of the same call, that is  $j = i + 1$  with  $j \in \mathcal{C} \setminus \mathcal{S}$ , or if  $j$  is the starting phase of the following idle period, that is  $\{i + 1, \dots, j - 1\} \subset \mathcal{C} \setminus \mathcal{S}$  and  $j \in \mathcal{I}$ . Similarly, for any idle phase  $i \in \mathcal{I}$ , we let  $p_{ij} = 0$  for all  $j$  except if  $j$  is the following phase of the same idle period, that is  $j = i + 1$  with  $j \in \mathcal{I}$ , or if  $j$  is the starting phase of the following call, that is  $\{i + 1, \dots, j - 1\} \subset \mathcal{I}$  and  $j \in \mathcal{C} \setminus \mathcal{S}$ . We have  $p_i = 0$  for all  $i$  except if  $i$  is a phase of the last call of a session, that is  $i \in \mathcal{C}$  and  $\{i + 1, \dots, j - 1\} \in \mathcal{C} \setminus \mathcal{S}$  for some  $j \in \mathcal{S}$  corresponding to the starting phase of another session.

For any call phase  $i \in \mathcal{C}$ , we refer to the smallest index  $j$  such that  $j \in \mathcal{C}$  and  $\{j + 1, \dots, i\} \subset \mathcal{C} \setminus \mathcal{S}$  as the starting phase of the call. Similarly, for any idle phase  $i \in \mathcal{I}$ , we refer to the smallest index  $j$  such that  $\{j, j + 1, \dots, i\} \subset \mathcal{I}$  as the starting phase of the idle period. For any call or idle phase  $i \in \mathcal{C} \cup \mathcal{I}$ , we refer to the largest index  $j \leq i$  such that  $j \in \mathcal{S}$  as the starting phase of the corresponding session. We denote this index by  $s(i)$ .

It is clear that this queueing network can represent any type of session, as illustrated by Figure 2. Sessions are mutually independent and the probability that a new session is of type  $i \in \mathcal{S}$  is given by:

$$\frac{\nu_i}{\nu}.$$

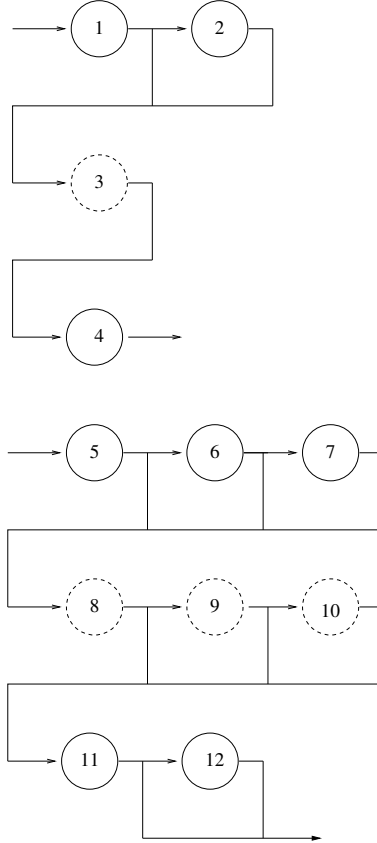


Figure 2: Two types of two-call sessions represented as sequences of exponential phases, with starting phases  $\mathcal{S} = \{1, 5\}$ , call phases  $\mathcal{C} = \{1, 2, 4, 5, 6, 7, 11, 12\}$  and idle phases  $\mathcal{I} = \{3, 8, 9, 10\}$  (dashed circles).

Provided the number of queues  $N$  is sufficiently large, we may represent any traffic characteristics, with arbitrary distributions for the number of calls per session, call durations and idle durations. There may be arbitrary correlation between these random variables. In the example of Figure 2 for instance, assuming all exponential phases have the same mean, large idle durations are typically preceded and followed by large call durations.

The overall traffic intensity, defined as the product of the call arrival rate and the mean call duration and usually expressed in Erlangs, is given by:

$$\rho = \sum_{i \in \mathcal{C}} \rho_i, \quad (5)$$

where  $\rho_i = \lambda_i/\mu_i$  denotes the traffic intensity at queue  $i$ . The arrival rate  $\lambda_i$  at any queue  $i$  that corresponds to the starting phase of a call or idle period is equal to  $\nu_s$ , with  $s = s(i)$ . For any other queue  $j$ , we have:

$$\lambda_j = \nu_s p_{i,i+1} p_{i+1,i+2} \cdots p_{j-1,j}, \quad (6)$$

where  $i$  denotes the starting phase of the corresponding call or idle period and  $s = s(i)$ .

### 3.2 Blocking probability

We now evaluate the blocking probability in the presence of  $C$  telephone lines. The set of admissible states is then:

$$\mathcal{A} = \left\{ x \in \mathbb{N}^N : \sum_{i \in \mathcal{C}} x_i \leq C \right\}.$$

Assume the retrial behavior is such that the conditions of Theorem 2 are satisfied. By the ‘Poisson Arrivals See Time Average’ property, the first call of any session sees the system in steady state. The probability such a call is blocked is equal to the stationary probability that all lines are occupied. Now in view of Corollary 1, customers leaving any *free* queue see the network in steady state. Noting that queues corresponding to idle phases are free, we deduce that the probability an arbitrary call is blocked is equal to the stationary probability that all lines are occupied.

For all  $n = 0, 1, \dots, C$ , the stationary probability that  $n$  lines are occupied is given by:

$$\bar{\pi}(n) = \sum_{x: \sum_{i \in \mathcal{C}} x_i = n} \pi'(x).$$

In view of Theorem 2, we get:

$$\bar{\pi}(n) = \bar{\pi}(0) \frac{\rho^n}{n!},$$

with

$$\bar{\pi}(0) = \left( \sum_{n=0}^C \frac{\rho^n}{n!} \right)^{-1},$$

where  $\rho$  denotes the overall traffic intensity, given by (5). We deduce the blocking probability of an arbitrary call:

$$B = \frac{\frac{\rho^C}{C!}}{\sum_{n=0}^C \frac{\rho^n}{n!}}. \quad (7)$$

This is the well-known Erlang formula. It does not depend on any traffic characteristics described in §3.1 beyond the traffic intensity  $\rho$ . The blocking probability is the same for all calls independently of the session they belong to and of their location in that session (e.g., the first, second or last call of the session).

## 4 Retrial behavior

We now describe two types of retrial behavior for which the conditions of Theorem 2 are satisfied. We first consider jump-over blocking then random retrials with fixed retrial probability  $r$ . Though the latter contains the former as a special case (for  $r = 0$ ), we present them as separate schemes for the sake of clarity.



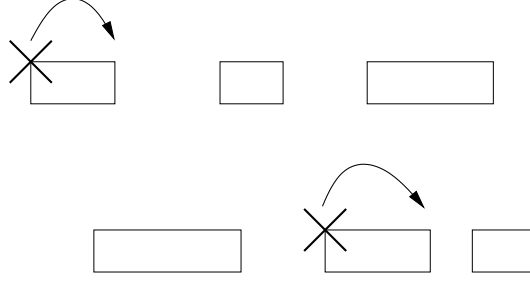


Figure 3: Jump-over blocking.

#### 4.1 Jump-over blocking

We first assume that sessions go on in case of blocking, as illustrated by Figure 3. If a call is blocked, the following idle period starts immediately. If the call is the last of the session, the session ends. We here assume that the retrial behavior is the same for all sessions. This assumption will be relaxed in §4.2 where the retrial probability may depend on the considered session.

Let  $x \in \mathcal{A}$  be an admissible state. For any starting phase of a session  $i \in \mathcal{S}$  such that  $x + e_i \notin \mathcal{A}$ , we have  $\nu'_i(x) = 0$ . If the session consists of more than one call, we also have  $\nu'_j(x) = \nu_i$ , where  $j$  denotes the starting phase of following idle period. Similarly, for any starting phase of a call  $i \in \mathcal{C} \setminus \mathcal{S}$  such that  $x + e_i \notin \mathcal{A}$ , we have  $p'_{ji}(x) = 0$  where  $j$  denotes any phase of the preceding idle period. If the call is not the last of the session, we also have  $p'_{jk}(x) = p_{ji}$  where  $k$  denotes the starting phase of the following idle period. Other external arrival rates and routing probabilities are the same as in the absence of call blocking, as described in §3.1.

Jump-over blocking is a standard rerouting scheme in queueing theory, see e.g. [5, 17]. In particular, it is known that the conditions of Theorem 2 are satisfied. We prove the result for the sake of completeness. For all  $x$  such that

$$\sum_{i \in \mathcal{C}} x_i < C,$$

the solution of the corresponding traffic equations (3) is unique and given by  $\lambda'_i(x) = \lambda_i$ . Now let  $x$  be such that:

$$\sum_{i \in \mathcal{C}} x_i = C.$$

We have  $\lambda'_i(x) = 0$  for all  $i \in \mathcal{C}$ . It remains to prove that  $\lambda'_i(x) = \lambda_i$  for all  $i \in \mathcal{I}$ . Let  $i, i+1, \dots, k$  be the successive phases of the same idle period. We denote by  $s = s(i)$  the starting phase of the corresponding session. The traffic equations are  $\lambda'_i(x) = \nu_s$  and

$$\lambda'_{j+1}(x) = \lambda'_j(x)p_{j,j+1}, \quad j = i, \dots, k-1.$$

It then follows from (6) that:

$$\lambda'_j(x) = \lambda_j, \quad j = i, \dots, k.$$

The conditions of Theorem 2 are satisfied.

#### 4.2 Random retrials

We now consider random retrials, as illustrated by Figure 4. Each blocked call is reattempted with a fixed probability  $r$  after an idle period of random duration. The session goes on as in the jump-over blocking model with probability  $1 - r$ . The duration of the idle period between two attempts has the same distribution as the idle period that precedes the considered call. If the considered call is the first of a session, there is no retrial and the session is lost with probability  $r$  and goes on with probability  $1 - r$ .

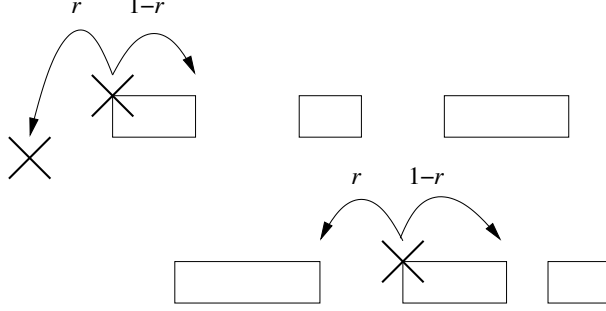


Figure 4: Blocking with random retrial.

Note that the retrial behavior corresponds to the jump-over blocking scheme when  $r = 0$ . If  $r = 1$ , all calls of the session are reattempted until they are accepted, except the first call for which there is no retrial. Thus in this case, either the first call is accepted and all calls of the session are eventually accepted, or the first call is rejected and the entire session is rejected. In the following, we allow the retrial probability to depend on the session type. For all  $i \in \mathcal{S}$ , we denote by  $r_i$  the retrial probability of the corresponding session.

Let  $x \in \mathcal{A}$  be an admissible state. For any starting phase of a session  $i \in \mathcal{S}$  such that  $x + e_i \notin \mathcal{A}$ , we have  $\nu'_i(x) = 0$ . If the session consists of more than one call, we also have  $\nu'_j(x) = \nu_i(1 - r_i)$  where  $j$  is the starting phase of the following idle period. Similarly, for any starting phase of a call  $i \in \mathcal{C} \setminus \mathcal{S}$  such that  $x + e_i \notin \mathcal{A}$ , we have  $p'_{ji}(x) = 0$  where  $j$  denotes any phase of the preceding idle period, and  $p'_{jk}(x) = p_{ji}r_s$  where  $k$  denotes the starting phase of this idle period and  $s = s(i)$  denotes the starting phase of the session. If this call is not the last of the session, we also have  $p'_{jl}(x) = p_{ji}(1 - r_s)$  where  $l$  denotes the starting phase of the following idle period. Other external arrival rates and routing probabilities are the same as in the absence of call blocking, as described in §3.1.

For all states  $x$  such that

$$\sum_{i \in \mathcal{C}} x_i < C,$$

the solution of the traffic equations (3) is unique and given by  $\lambda'_i(x) = \lambda_i$ .

Now let  $x$  be such that:

$$\sum_{i \in \mathcal{C}} x_i = C.$$

Again, we have  $\lambda'_i(x) = 0$  for all  $i \in \mathcal{C}$ . It remains to prove that  $\lambda'_i(x) = \lambda_i$  for all  $i \in \mathcal{I}$ . Let  $i, i+1, \dots, k$  be the successive phases of the same idle period. We denote by  $s = s(i)$  the starting phase of the corresponding session. The traffic equations are:

$$\lambda'_i(x) = \nu_s(1 - r_s) + \sum_{j=i}^k \lambda'_j(x) p_{j,k+1} r_s,$$

and

$$\lambda'_{j+1}(x) = \lambda'_j(x) p_{j,j+1}, \quad j = i, \dots, k-1.$$

Using the fact that  $p_{k,k+1} = 1$  and  $p_{j,j+1} + p_{j,k+1} = 1$  for all  $j = i, \dots, k-1$ , we deduce:

$$\lambda'_i(x) = \nu_s(1 - r_s) + \lambda'_i(x) r_s,$$

If  $r_s < 1$ , the solution to this equation is unique, given by  $\lambda'_i(x) = \nu_s$ . If  $r_s = 1$ , the subnetwork corresponding to session  $s$  is a set of closed queueing networks in state  $x$  (one per idle period). The arrival rates are defined up to a multiplicative constant and we choose  $\lambda'_i(x) = \nu_s$ . In both cases, we obtain:

$$\lambda'_j(x) = \lambda_j, \quad j = i, \dots, k.$$

The conditions of Theorem 2 are satisfied.

It is worth noting that the results extend to the case where a blocked call is reattempted with a fixed probability after an idle period of *constant* duration equal to the preceding idle period, instead of a random duration with the same distribution. By increasing the number of queues  $N$ , the distribution of call durations and idle durations may indeed be made as close to deterministic as desired. In the limit, the random nature of traffic reduces to the Poisson session arrivals and the probability  $\nu_i/\nu$  a session is of type  $i \in \mathcal{S}$ .

## 5 Extensions

We show in this section that the insensitivity property of the Erlang formula extends to more general loss networks and to the Engset formula under the same conditions on the retrial behavior as described in the previous section.

### 5.1 Multi-rate systems

We first consider multi-rate systems that have been extensively studied in the 70's and 80's [7, 10, 11, 13, 15] and are still used to dimension wired and wireless networks when voice, data, audio and video flows are multiplexed within constant bit rate circuits. There are  $K$  call classes. Class- $k$  calls arrive as a Poisson process and require a circuit of  $c_k$  bit/s. We denote by  $\rho_k$  the traffic intensity of class- $k$  calls in Erlangs.

Denoting by  $C$  the link capacity in bit/s, it is well known that the stationary probability that  $n_k$  class- $k$  calls are in progress,  $k = 1, \dots, K$ , is independent of the call duration distributions and given by:

$$\bar{\pi}(n) = \bar{\pi}(0) \prod_{k=1}^K \frac{\rho_k^{n_k}}{n_k!}, \quad n.c \leq C$$

with

$$\bar{\pi}(0) = \left( \sum_{n:n.c \leq C} \prod_{k=1}^K \frac{\rho_k^{n_k}}{n_k!} \right)^{-1},$$

where  $n.c$  denotes the scalar product:

$$n.c \stackrel{\text{def}}{=} \sum_{k=1}^K n_k c_k.$$

The blocking probability of class- $k$  calls is then given by:

$$B_k = \sum_{n:C-c_k < n.c \leq C} \bar{\pi}(n).$$

In view of Theorem 2, this result still holds under the more general assumption of Poisson *session* arrivals provided the retrial behavior is that described in Section 4.

Specifically, consider a partition of  $K$  subsets  $\mathcal{C}_1, \dots, \mathcal{C}_K$  of the set of queues  $\mathcal{C}$ , corresponding to calls of class 1,  $\dots$ ,  $K$ , respectively. The set of admissible states is:

$$\mathcal{A} = \left\{ x \in \mathbb{N}^N : \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} x_i c_k \leq C \right\}.$$

The results of Sections 3-4 readily apply provided sessions are *homogeneous* in the sense that all calls belonging to the same session are of the same class. We give an example in Section 6 showing the sensitivity of the blocking probability when sessions are composed of calls of different classes.

## 5.2 Loss networks

We now consider a general model with several resources. Specifically, the network consists of  $L$  links of respective capacities  $C_1, \dots, C_L$  in bit/s. There are  $K$  call classes. Class- $k$  calls arrive as a Poisson process and require a circuit of  $c_k$  bit/s through links  $a_k \subset \{1, \dots, L\}$ . We still denote by  $\rho_k$  the traffic intensity of class- $k$  calls in Erlangs.

Again, the stationary probability that  $n_k$  class- $k$  calls are in progress,  $k = 1, \dots, K$ , is independent of the call duration distributions and given by [1, 4, 12]:

$$\bar{\pi}(n) = \bar{\pi}(0) \prod_{k=1}^K \frac{\rho_k^{n_k}}{n_k!}, \quad An \leq C$$

with

$$\bar{\pi}(0) = \left( \sum_{n: An \leq C} \prod_{k=1}^K \frac{\rho_k^{n_k}}{n_k!} \right)^{-1},$$

where  $A$  is the  $L \times K$ -dimensional matrix whose  $l, k$  entry is equal to  $c_k$  if  $l \in a_k$  and to 0 otherwise,  $C$  is the  $L$ -dimensional vector whose  $l$ -th component is equal to  $C_l$  and  $\leq$  denotes the component-wise order.

The blocking probability of class- $k$  calls is then given by:

$$B_k = \sum_{n: C - Ae_k < An \leq C} \bar{\pi}(n),$$

where  $e_k$  denotes the  $K$ -dimensional unit vector with 1 in component  $k$  and 0 elsewhere and for all vectors  $x, y$ ,  $x < y$  means  $x \leq y$  component-wise and  $x \neq y$ . Again, the result remains valid for Poisson session arrivals provided sessions are homogeneous and the retrial behavior is that described in Section 4.

## 5.3 The Engset formula

Finally, we consider the Engset model with a fixed number  $M$  of permanent sessions. Though the results apply to general loss networks as well, we focus on the simple example of a single link of  $C$  telephone lines as originally considered by Engset [6]. We make the natural assumption that  $M > C$ .

Let  $\rho$  be the ratio of the mean call duration to the mean idle duration. Under i.i.d. assumptions, it is well known that the stationary probability that  $n$  calls are in progress is independent of the distributions of call durations and idle durations beyond  $\rho$  and given by:

$$\bar{\pi}(n) = \bar{\pi}(0) \binom{M}{n} \rho^n, \quad n \leq C \tag{8}$$

with

$$\bar{\pi}(0) = \left( \sum_{n=0}^C \binom{M}{n} \rho^n \right)^{-1}.$$

In addition, new calls see a system of  $M - 1$  permanent sessions in steady state when they arrive.

We deduce the call blocking probability:

$$B = \frac{\binom{M-1}{C} \rho^C}{\sum_{n=0}^C \binom{M-1}{n} \rho^n}. \tag{9}$$

This is the Engset formula. We now prove that the formula is valid for general traffic characteristics including correlated call durations and idle durations provided the retrial behavior is that described in Section 4.

We consider a closed network of  $N$  infinite-server queues and  $M$  customers. Service times at queue  $i$  are independent, exponentially distributed of mean  $1/\mu_i$ , for some  $\mu_i > 0$ . Once served at queue  $i$ , customers move to queue  $j$  with probability  $p_{ij}$ , with:

$$\sum_{j=1}^N p_{ij} = 1.$$

Routing is irreducible in the sense that each customer visits all queues. The arrival rate  $\lambda_i$  at queue  $i$  is uniquely defined, up to a multiplicative constant, by the traffic equations:

$$\lambda_i = \sum_{j=1}^N \lambda_j p_{ji}. \quad (10)$$

We define the traffic intensity at queue  $i$  by  $\rho_i = \lambda_i/\mu_i$ .

The stationary distribution of the Markov process  $X(t)$  describing the network state is given by:

$$\pi(x) = G \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!}, \quad \sum_{i=1}^N x_i = M, \quad (11)$$

where  $G$  is the normalizing constant, given by:

$$G = \frac{M!}{\left(\sum_{i=1}^N \rho_i\right)^M}.$$

We denote by  $\mathcal{C}$  the set of queues corresponding to call phases and by  $\mathcal{I}$  the set of queues corresponding to idle phases, with  $\mathcal{C} \cup \mathcal{I} = \{1, \dots, N\}$ . As explained in §3.1, the above queueing network may represent any traffic characteristics provided the number of queues  $N$  is sufficiently large. The ratio of the mean call duration to the mean idle duration is given by:

$$\varrho = \frac{\sum_{i \in \mathcal{C}} \rho_i}{\sum_{i \in \mathcal{I}} \rho_i}.$$

In view of (11), the stationary probability that  $n$  lines are occupied is given by:

$$\bar{\pi}(n) = \sum_{x: \sum_{i \in \mathcal{C}} x_i = n} \pi(x) = \bar{\pi}(0) \binom{M}{n} \varrho^n. \quad (12)$$

Again, we assume that queues are numbered in such a way that phases of successive calls and idle periods correspond to increasing indices. By convention, we let  $1 \in \mathcal{C}$ ,  $N \in \mathcal{I}$  and  $p_{N1} = 1$ . For any call phase  $i \in \mathcal{C}$ , we have  $p_{i,i+1} + p_{ij} = 1$  where  $j$  is the starting phase of the following idle period. For any idle phase  $i \in \mathcal{I}$ ,  $i \neq N$ , we have  $p_{i,i+1} + p_{ij} = 1$  where  $j$  is the starting phase of the following call. The starting phases of calls and idle periods have the same arrival rate. Since arrival rates are defined up to a multiplicative constant, we let  $\lambda_i = 1$  for the corresponding queues  $i$ . For any other queue  $j$ , we have:

$$\lambda_j = p_{i,i+1} p_{i+1,i+2} \dots p_{j-1,j}. \quad (13)$$

Now consider the restricted state space:

$$\mathcal{A} = \left\{ x \in \mathbb{N}^N : \sum_{i=1}^N x_i = M, \sum_{i \in \mathcal{C}} x_i \leq C \right\}.$$

Denote by  $p'_{ij}(x)$  the new routing probability from queue  $i$  to queue  $j$  in state  $x$ , with  $p'_{ij}(x) = 0$  if  $x - e_i + e_j \notin \mathcal{A}$ . Let  $\lambda'_i(x)$  be the (relative) arrival rate of customers to queue  $i$  in state  $x$ . The corresponding traffic equations are:

$$\lambda'_i(x) = \sum_{j=1}^N \lambda'_j(x) p'_{ji}(x + e_j). \quad (14)$$

Let  $X'(t)$  be the Markov process describing the network state. We have the analog of Theorem 2:

**Theorem 3** *If the arrival rates defined by:*

$$\lambda'_i(x) = \lambda_i \text{ if } x + e_i \in \mathcal{A}, \quad \lambda'_i(x) = 0 \text{ otherwise}$$

*are a solution to the traffic equations (14) for all queues  $i$  and all states  $x \in \mathbb{N}^N$  such that  $\sum_{i=1}^N x_i = M - 1$ , then the stationary distribution  $\pi'$  of the Markov process  $X'(t)$  is the restriction of  $\pi$  to the set of admissible states  $\mathcal{A}$ , that is:*

$$\pi'(x) = G' \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \text{ if } x \in \mathcal{A}, \quad \pi'(x) = 0 \text{ otherwise,}$$

with

$$G' = \left( \sum_{x \in \mathcal{A}} \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \right)^{-1}.$$

*Proof.* It follows from the traffic equations (14) that for all queues  $i$  and all states  $x$  such that  $\sum_{i=1}^N x_i = M - 1$ ,

$$\pi'(x + e_i) \mu_i (x_i + 1) = \sum_{j=1}^N \pi'(x + e_j) \mu_j (x_j + 1) p'_{ji}(x + e_j).$$

The equilibrium equations follow by summation:

$$\pi'(x) \sum_{i=1}^N \mu_i x_i = \sum_{i,j=1}^N \pi'(x + e_j - e_i) \mu_j (x_j + 1) p'_{ji}(x + e_j - e_i),$$

where we use the convention that  $\pi'(x) = 0$  if  $x \notin \mathbb{N}^N$ . □

The corresponding ‘Departures See Time Average’ property is the following. Let:

$$\mathcal{A}' = \left\{ x \in \mathbb{N}^N : \sum_{i=1}^N x_i = M - 1, \sum_{i \in \mathcal{C}} x_i \leq C \right\}.$$

A queue  $i$  is said to be *free* if  $x + e_i \in \mathcal{A}$  for all states  $x \in \mathcal{A}'$ . Note that the set of free queues here corresponds to the set of idle phases  $\mathcal{I}$ .

**Corollary 2** *Under the conditions of Theorem 3, customers leaving any free queue see a network with  $M - 1$  customers in steady state immediately after their departure.*

*Proof.* The probability  $\pi'_i(x)$  that customers leaving free queue  $i$  see the network in state  $x$  immediately after their departure is proportional to  $\pi'(x + e_i) \mu_i (x_i + 1)$ , so that:

$$\pi'_i(x) = \frac{\pi'(x + e_i) \mu_i (x_i + 1)}{\sum_{y \in \mathcal{A}'} \pi'(y + e_i) \mu_i (y_i + 1)}.$$

In view of Theorem 3, we have:

$$\forall x \in \mathcal{A}', \quad \pi'(x + e_i)\mu_i(x_i + 1) = \frac{G'}{G''}\pi''(x)\lambda_i,$$

where  $\pi''$  denote the stationary distribution of the network with  $M - 1$  customers:

$$\pi''(x) = G'' \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \text{ if } x \in \mathcal{A}', \quad \pi''(x) = 0 \text{ otherwise,}$$

with

$$G'' = \left( \sum_{x \in \mathcal{A}'} \prod_{i=1}^N \frac{\rho_i^{x_i}}{x_i!} \right)^{-1}.$$

We deduce:

$$\pi'_i(x) = \frac{\pi''(x)\lambda_i}{\sum_{y \in \mathcal{A}'} \pi''(y)\lambda_i} = \pi''(x).$$

□

We verify as in Section 4 that the conditions of Theorem 3 are satisfied for both jump-over blocking and random retrials. It then follows from Corollary 2 that the blocking probability is equal to the probability that all lines are occupied in the presence of  $M - 1$  customers, independently of the considered call. This blocking probability is given by the Engset formula (9).

## 6 Examples

Finally, we illustrate the above insensitivity results by some examples.

### 6.1 Burst arrivals within sessions

The Erlang model with sessions may include call bursts, that is successions of calls separated by relatively short idle periods, as illustrated by Figure 5. It is then somewhat surprising that the second call of a session is blocked with the same probability as the first call of the session. Intuitively, given the fact that the first call was accepted, the second call is also accepted with a high probability after a short idle period.

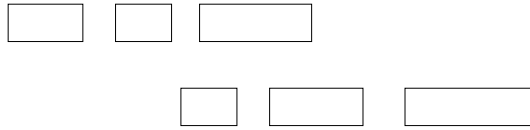


Figure 5: Burst arrivals within sessions.

Consider for instance two-call sessions with exponential holding times of mean  $1/\mu$  and exponential idle periods of mean  $\tau$ , for some  $\mu, \tau > 0$ . Sessions arrive according to a Poisson process of intensity  $\lambda$ . In view of Theorem 2, the number of idle sessions is independent of the number of ongoing calls and has a Poisson distribution of mean  $\lambda\tau$ . In particular, the probability that there is no other idle session at the end of the first call is equal to  $e^{-\lambda\tau}$ . Moreover, the probability that no other session is generated during an idle period is equal to:

$$\frac{1}{1 + \lambda\tau}.$$

We conclude that the probability that the second call of the session is accepted given the fact that the first call was accepted indeed tends to 1 when the mean idle period  $\tau$  tends to 0.

How to solve the paradox? In case of jump-over blocking, it is sufficient to observe that the probability that the second call of the session is blocked given the fact that the first call was blocked also tends to 1 when the mean idle period  $\tau$  tends to 0. This follows from the fact that the probability that no ongoing call ends during an idle period is equal to:

$$\frac{1}{1 + C\mu\tau},$$

where  $C$  denotes the number of lines. In the limit  $\tau \rightarrow 0$ , the blocking probability of the second call of the session is *equal* to the blocking probability of the first call of the session, which is given by the Erlang formula.

For random retrials with retrial probability  $r < 1$ , a fraction  $1 - r$  of the first calls blocked generate a second call. Again, the successive attempts of this second call will be blocked with high probability for a short idle period. In the limit  $\tau \rightarrow 0$ , all these attempts are blocked. Since the average number of attempts is equal to  $(1 - r)^{-1}$ , the average number of second calls blocked per first call blocked is equal to 1. The blocking probability of the second call is equal to the blocking probability of the first call of the session, which is given by the Erlang formula.

Finally, consider the case where the retrial probability is  $r = 1$ . The entire session is then blocked if the first call of the session is blocked (cf. Figure 4). In this case, we solve the paradox as follows: even if the probability that the second call of the session is accepted conditionally to the fact that the first call was accepted is close to 1 for a short idle period, the blocking probability is *not* equal to 0. And if the first attempt of second call is blocked, next reattempts will also be blocked with a high probability. In the limit  $\tau \rightarrow 0$ , the blocking probability of the second call is equal to 0 but if blocked, the number of reattempts is infinite. The blocking probability is in fact equal to the Erlang formula independently of the mean idle duration  $\tau$ .

## 6.2 Burst arrivals of calls

It is worth observing that, even if a session may generate call bursts, each call (except the first) starts after the previous call of the same session ends. This makes a significant difference with more artificial traffic models where non-Poisson call arrivals are generated based on statistics on observed call interarrival times [14]. In this case, the blocking probability may indeed be strongly sensitive to traffic characteristics, see e.g. [20]. We illustrate this sensitivity on the simple example illustrated by Figure 6 where calls arrive in bursts of fixed size  $b$  according to a Poisson process of intensity  $\lambda$ .

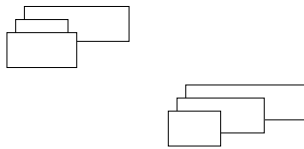


Figure 6: Burst arrivals of calls.

Let  $C$  be the number of telephone lines, with  $C > b$ . Assume the holding times are independent, exponentially distributed of mean  $1/\mu$ , for some  $\mu > 0$ .

The stationary probability  $\bar{\pi}$  of the number of ongoing calls satisfies the equilibrium equations:

$$(\lambda + n\mu)\bar{\pi}(n) = (n + 1)\mu\bar{\pi}(n + 1)$$

for all  $n = 0, \dots, b - 1$ , and

$$(\lambda + n\mu)\bar{\pi}(n) = (n + 1)\mu\bar{\pi}(n + 1) + \lambda\bar{\pi}(n - b)$$



for all  $n = b, \dots, C - 1$ . The blocking probability is then:

$$B = \sum_{n=0}^{b-1} \frac{b-n}{b} \bar{\pi}(C-n).$$

Figure 7 gives the blocking probability with respect to the traffic intensity  $b \times \lambda/\mu$  for  $C = 50$  telephone lines. We observe that the burst size  $b$  has a strong impact on the blocking probability. Note that a conservative estimate of the blocking probability is given by the corresponding Erlang formula where each burst of  $b$  calls is considered as a single call holding  $b$  lines simultaneously during a mean time  $1/\mu$  (cf. the figure, the estimate is exact for  $b = 1$ ). The result is indeed very sensitive to the burst size  $b$ .

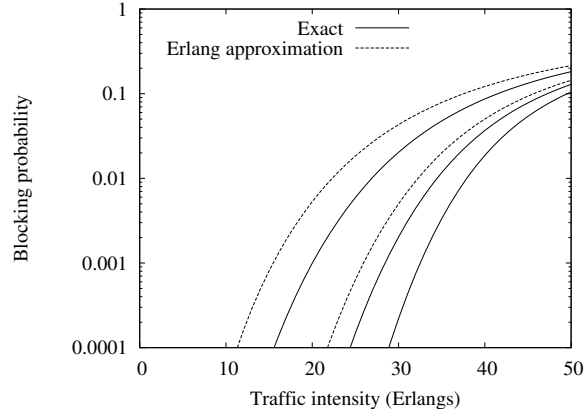


Figure 7: Blocking probability for various values of the burst size ( $b = 1, 2, 5$ , from bottom to top).

### 6.3 Impact of retrial behavior

We now illustrate the impact of retrial behavior. We focus on the first call of a session, for which there is no retrial in both schemes described in Section 4. In the simple case of Poisson call arrivals with retrial, the model reduces to a classical retrial queue, for which a rich literature exists, see [9, 19]. Such queues are known to be sensitive.

Consider for instance the Erlang model with  $C$  telephone lines, call arrival rate  $\lambda$  and mean holding time  $1/\mu$ , for some  $\mu > 0$ . We denote by  $\rho = \lambda/\mu$  the corresponding traffic intensity in Erlangs. Assume that each call is reattempted until it is accepted. Successive attempts are separated by independent, exponentially distributed idle periods of mean  $\tau$ . The number of idle sessions then increases continuously if  $\rho > C$ . If  $\rho < C$ , the number of idle sessions remains stable but the blocking probability of successive attempts is sensitive to the mean idle period  $\tau$ .

In the limiting case  $\tau \rightarrow 0$  for instance, an idle session becomes active as soon as a telephone line becomes idle. The blocking probability is then given by the corresponding Erlang C formula [8]:

$$B = \frac{\frac{\rho^C}{C!}}{\frac{C-\rho}{C} \sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{C!}}.$$

In the other limiting case  $\tau \rightarrow \infty$ , the overall call arrival process (including reattempts) tends to a Poisson process of intensity  $\lambda/(1-B)$ , where  $B$  denotes the corresponding blocking probability. The blocking

probability can then be determined by the fixed point equation:

$$B = E_B\left(\frac{\rho}{1-B}\right),$$

where  $E_B(\alpha)$  denotes the Erlang formula (also referred to as the Erlang B formula) for a traffic intensity  $\alpha$ :

$$E_B(\alpha) = \frac{\frac{\alpha^C}{C!}}{\sum_{n=0}^C \frac{\alpha^n}{n!}}.$$

The results are illustrated by Figure 8 for  $C = 10$  telephone lines.

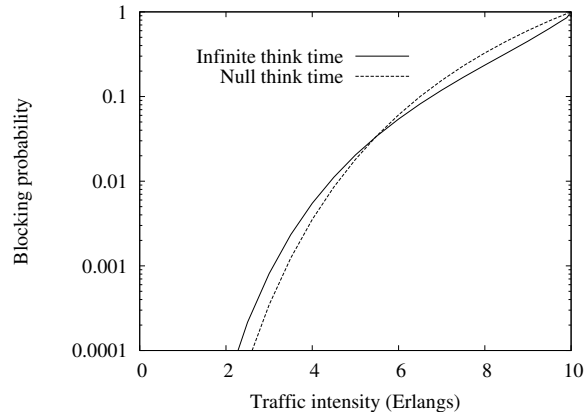


Figure 8: Sensitivity of the blocking probability to the mean idle period between successive attempts.

## 6.4 Loss networks with non-homogeneous sessions

We now show that successive calls of the same session *must* require the same network resources to preserve insensitivity. Consider for instance a multi-rate system as described in §5.1 with  $K = 2$  call classes and  $c_1 < c_2$ . Sessions arrive as a Poisson process and consist of two calls, a class-1 call followed by a class-2 call after a random idle period. Blocked calls are reattempted with probability  $r$  as described in Section 4. Such a system is sensitive. In the limiting case  $c_1 = 0$ , first calls are always accepted and the system behaves for second calls as a retrial queue, cf. §6.3. In the particular case where  $r = 1$ ,  $c_2 = 1$  and  $C = 10$ , the sensitivity of the blocking probability to the idle duration is given by Figure 8.

## 6.5 Multi-class Engset formula

Finally, we consider the Engset model with  $K$  session classes. All sessions of the same class have the same traffic characteristics. Let  $\varrho_k$  be the ratio of the mean holding time to the mean idle period for class- $k$  sessions. There are  $M_k$  such sessions. The stationary probability that  $n_k$  class- $k$  calls are in progress,  $k = 1, \dots, K$ , is independent of the holding time and idle period distributions and given by:

$$\bar{\pi}(n) = \bar{\pi}(0) \prod_{k=1}^K \binom{M_k}{n_k} \varrho_k^{n_k}, \quad n_k \leq M_k, \quad \sum_{k=1}^K n_k \leq C$$

with

$$\bar{\pi}(0) = \left( \sum_{n: n_k \leq M_k, \sum_{k=1}^K n_k \leq C} \prod_{k=1}^K \binom{M_k}{n_k} \varrho_k^{n_k} \right)^{-1}.$$

In addition, new class- $k$  calls see a system with  $M_k - 1$  class- $k$  sessions and  $M_l$  class- $l$  sessions,  $l \neq k$ , in steady state when they arrive. We deduce that the blocking probability depends the session class. As above, the corresponding multi-class Erlang formula is valid for general traffic characteristics including correlated call durations and idle durations provided the retrial behavior is that described in Section 4.

The sensitivity to the traffic mix is illustrated on Figure 9 for  $C = 10$  telephone lines and  $K = 2$  classes, with  $M_1 = 20$ ,  $M_2 = 1$  and  $\varrho_2 \rightarrow \infty$ . The class-2 session is either active or in continuous retrial if blocked. The class-1 blocking probability is given by the Engset formula for  $C - 1$  telephone lines and  $M_1$  class-1 sessions:

$$B_1 = \frac{\binom{M_1 - 1}{C - 1} \varrho_1^{C-1}}{\sum_{n=0}^{C-1} \binom{M_1 - 1}{n} \varrho_1^n}.$$

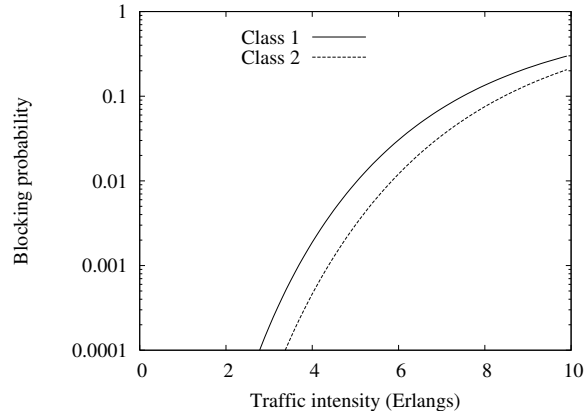


Figure 9: Sensitivity of the blocking probability to the session class.

The class-2 blocking probability is given by the Engset formula for  $C$  telephone lines and  $M_1 + 1$  class-1 sessions:

$$B_2 = \frac{\binom{M_1}{C} \varrho_1^C}{\sum_{n=0}^C \binom{M_1}{n} \varrho_1^n}.$$

It is worth noting that the average blocking probability depends on the traffic mix through the number of sessions of each class,  $M_1, \dots, M_K$ , and the corresponding ratios of the mean call duration to the mean idle duration,  $\varrho_1, \dots, \varrho_K$ . Engset conjectured in his 1915 report that the worst case is obtained in the homogeneous case where  $\varrho_1 = \dots = \varrho_K$  [6]. This result was proved by Dartois much later [3].

## 7 Conclusion

The multiplexing of various traffic streams like voice, data, audio and video flows may generate non-Poisson flow arrivals. This raises the issue of the applicability of standard teletraffic models like the Erlang model

when these flows are multiplexed in physical or virtual constant bit rate circuits. We have demonstrated that these models remain applicable provided calls are generated within independent sessions, which is typically the case in practice. These sessions may arrive as a Poisson process as in the Erlang model or be permanent as in the Engset model, covering the whole range of traffic conditions. Thus these models satisfy even stronger insensitivity properties than generally believed.

A critical role is played by the retrial behavior. We have described two schemes, namely jump-over blocking and random retrials, for which the insensitivity property is indeed satisfied. We have also given examples of other schemes for which the insensitivity property is violated. A natural question of interest is to find the whole set of schemes for which the insensitivity property holds. This will be the subject of future research.

## References

- [1] D.Y. Burman, J.P. Lehoczky and Y. Lim, Insensitivity of blocking probabilities in a circuit-switching network, *J. Appl. Probab.* 21 (1984) 850–859.
- [2] J.W. Cohen, The Generalized Engset Formula, *Phillips Telecommunications Review* 18 (1957) 158–170.
- [3] J.P. Dartois, Lost call cleared systems with unbalanced traffic sources, in: *Proc. of 6th International Teletraffic Congress*, 1970.
- [4] Z. Dziong, J.W. Roberts, Congestion probabilities in a circuit-switched integrated services network, *Performance Evaluation* 7-4 (1987) 267–284.
- [5] A. Economou and D. Fakinos, Product form stationary distributions for queueing networks with blocking and rerouting, *Queueing Systems* 30 (1998) 251–260.
- [6] T.O. Engset, On the calculation of switches in an automatic telephone system, in: *Tore Olaus Engset: The man behind the formula*, Eds: A. Myskjå, O. Espvik, 1998. First appeared as an unpublished report in Norwegian, 1915.
- [7] O. Enomoto, H. Miyamoto, An Analysis of mixtures of multiple bandwidth traffic on time division switching networks, in: *Proc. of the 7th International Teletraffic Congress*, 1973.
- [8] A.K. Erlang, Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges, in: *The life and works of A.K. Erlang*, Eds: E. Brockmeyer, H.L. Halstrom, A. Jensen, 1948. First published in Danish, 1917.
- [9] G.I. Falin and J.G.C. Templeton, *Retrial Queues*, Chapman and Hall, 1997.
- [10] L.A. Gimpelson, Analysis of mixtures of wide and narrow-band traffic, in: *IEEE Trans. Comm. Technology* 13-3 (1965) 258–266.
- [11] J. S. Kaufman, Blocking in a shared resource environment, *IEEE Trans. Commun.* 29 (1981) 1474–1481.
- [12] F.P. Kelly, Loss networks, *Annals of Applied Probability* 1 (1991) 319–378.
- [13] F.P. Kelly, Blocking probabilities in large circuit-switched networks, *Advances in Applied Probability* 18 (1986) 473–505.
- [14] A. Pattavina, A. Parini, Modelling voice call interarrival and holding time distributions in mobile networks, in: *Proc. of 19th International Teletraffic Congress 19*, 2005.
- [15] J. W. Roberts, A service system with heterogeneous user requirement, in: *Performance of Data Communications Systems and Their Applications*, G. Pujolle, Ed. Amsterdam, The Netherlands: North-Holland, 1981, pp. 423–431.

- [16] K. Ross, *Multiservice Loss Networks for Broadband Telecommunications Networks*, Springer-Verlag, 1995
- [17] R.F. Serfozo, *Introduction to Stochastic Networks*, Springer Verlag, 1999.
- [18] B.A. Sevastyanov, An ergodic theorem for Markov processes and its application to telephone systems with refusals, *Theor. Probability Appl.* 2 (1957) 104–112.
- [19] R.W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, 1989.
- [20] Y. Wu, C. Williamson, Impacts of data call characteristics on multi-service CDMA system capacity, in: *Proc. of Performance*, 2005.