

From:

Wireless Communications: Principles and Practice

by

Theodore S. Rappaport

IEEE Press (The Institute of Electrical And Electronics Engineers, Inc.

Prentice Hall PTR

ISBN: 0-7803-1167-1

Copyright 1996

2.6 Trunking and Grade of Service

Cellular radio systems rely on *trunking* to accommodate a large number of users in a limited radio spectrum. The concept of trunking allows a large number of users to share the relatively small number of channels in a cell by providing access to each user, on demand, from a pool of available channels. In a trunked radio system, each user is allocated a channel on a per call basis, and upon termination of the call, the previously occupied channel is immediately returned to the pool of available channels.

Trunking exploits the statistical behavior of users so that a fixed number of channels or circuits may accommodate a large, random user community. The telephone company uses trunking theory to determine the number of telephone

circuits that need to be allocated for office buildings with hundreds of telephones, and this same principle is used in designing cellular radio systems. There is a trade-off between the number of available telephone circuits and the likelihood of a particular user finding that no circuits are available during the peak calling time. As the number of phone lines decreases, it becomes more likely that all circuits will be busy for a particular user. In a trunked mobile radio system, when a particular user requests service and all of the radio channels are already in use, the user is blocked, or denied access to the system. In some systems, a queue may be used to hold the requesting users until a channel becomes available.

To design trunked radio systems that can handle a specific capacity at a specific "grade of service", it is essential to understand trunking theory and queuing theory. The fundamentals of trunking theory were developed by Erlang, a Danish mathematician who, in the late 19th century, embarked on the study of how a large population could be accommodated by a limited number of servers [Bou88]. Today, the measure of traffic intensity bears his name. One Erlang represents the amount of traffic intensity carried by a channel that is completely occupied (i.e. 1 call-hour per hour or 1 call-minute per minute). For example, a radio channel that is occupied for thirty minutes during an hour carries 0.5 Erlangs of traffic.

The grade of service (GOS) is a measure of the ability of a user to access a trunked system during the busiest hour. The busy hour is based upon customer demand at the busiest hour during a week, month, or year. The busy hours for cellular radio systems typically occur during rush hours, between 4 p.m. and 6 p.m. on a Thursday or Friday evening. The grade of service is a benchmark used to define the desired performance of a particular trunked system by specifying a desired likelihood of a user obtaining channel access given a specific number of channels available in the system. It is the wireless designer's job to estimate the maximum required capacity and to allocate the proper number of channels in order to meet the *GOS*. *GOS* is typically given as the likelihood that a call is blocked, or the likelihood of a call experiencing a delay greater than a certain queuing time.

A number of definitions listed in Table 2.3 are used in trunking theory to make capacity estimates in trunked systems.

The traffic intensity offered by each user is equal to the call request rate multiplied by the holding time. That is, each user generates a traffic intensity of A_u Erlangs given by

$$A_u = \mu H \quad (2.13)$$

where H is the average duration of a call and μ is the average number of call requests per unit time. For a system containing U users and an unspecified number of channels, the total offered traffic intensity A , is given as

$$A = UA_U \quad (2.14)$$

Table 2.3 Definitions of Common Terms Used in Trunking Theory

<p><i>Set-up Time:</i> The time required to allocate a trunked radio channel to a requesting user.</p> <p><i>Blocked Call:</i> Call which cannot be completed at time of request, due to congestion. Also referred to as a <i>lost call</i>.</p> <p><i>Holding Time:</i> Average duration of a typical call. Denoted by H (in seconds).</p> <p><i>Traffic Intensity:</i> Measure of channel time utilization, which is the average channel occupancy measured in Erlangs. This is a dimensionless quantity and may be used to measure the time utilization of single or multiple channels. Denoted by A.</p> <p><i>Load:</i> Traffic intensity across the entire trunked radio system, measured in Erlangs.</p> <p><i>Grade of Service (GOS):</i> A measure of congestion which is specified as the probability of a call being blocked (for Erlang B), or the probability of a call being delayed beyond a certain amount of time (for Erlang C).</p> <p><i>Request Rate:</i> The average number of call requests per unit time. Denoted by n seconds⁻¹.</p>

Furthermore, in a C channel trunked system, if the traffic is equally distributed among the channels, then the traffic intensity per channel, A_c , is given as

$$A_c = UA_u/C \quad (2.15)$$

Note that the offered traffic is not necessarily the traffic which is *carried* by the trunked system, only that which is *offered* to the trunked system. When the

offered traffic exceeds the maximum capacity of the system, the carried traffic becomes limited due to the limited capacity (i.e. limited number of channels). The maximum possible carried traffic is the total number of channels, C , in Erlangs. The AMPS cellular system is designed for a GOS of 2% blocking. This implies that the channel allocations for cell sites are designed so that 2 out of 100 calls will be blocked due to channel occupancy during the busiest hour.

There are two types of trunked systems which are commonly used. The first type offers no queuing for call requests. That is, for every user who requests service, it is assumed there is no setup time and the user is given immediate access to a channel if one is available. If no channels are available, the requesting user is blocked without access and is free to try again later. This type of trunking is called *blocked calls cleared* and assumes that calls arrive as determined by a Poisson distribution. Furthermore, it is assumed that there are an infinite number of users as well as the following: (a) there are memoryless arrivals of requests, implying that all users, including blocked users, may request a channel at any time; (b) the probability of a user occupying a channel is exponentially distributed, so that longer calls are less likely to occur as described by an exponential distribution; and (c) there are a finite number of channels available in the trunking pool. This is known as an M/M/m queue, and leads to the derivation of the Erlang B formula (also known as the *blocked calls cleared* formula). The Erlang B formula determines the probability that a call is blocked and is a measure of the GOS for a trunked system which provides no queuing for blocked calls. The Erlang B formula is derived in Appendix A and is given by

$$Pr[\text{blocking}] = \frac{\frac{A^C}{C!}}{\sum_{k=0}^C \frac{A^k}{k!}} = GOS \quad (2.16)$$

where C is the number of trunked channels offered by a trunked radio system and A is the total offered traffic. While it is possible to model trunked systems with finite users, the resulting expressions are much more complicated than the Erlang B result, and the added complexity is not warranted for typical trunked systems which have users that outnumber available channels by orders of magnitude. Furthermore, the Erlang B formula provides a conservative estimate of the GOS, as the finite user results always predict a smaller likelihood of blocking. The capacity of a trunked radio system where blocked calls are lost is tabulated for various values of GOS and numbers of channels in Table 2.4.

Table 2.4 Capacity of an Erlang B System

Number of Channels C	Capacity (Erlangs) for GOS			
	= 0.01	= 0.005	= 0.002	= 0.001
2	0.153	0.105	0.065	0.046
4	0.869	0.701	0.535	0.439
5	1.36	1.13	0.900	0.762
10	4.46	3.96	3.43	3.09
20	13.0	11.1	10.1	9.11
24	15.3	14.2	13.0	12.2
40	29.0	27.3	25.7	24.5
70	56.1	53.7	51.0	49.2
100	84.1	80.9	77.4	75.2

The second kind of trunked system is one in which a queue is provided to hold calls which are blocked. If a channel is not available immediately, the call request may be delayed until a channel becomes available. This type of trunking is called *Blocked Calls Delayed*, and its measure of GOS is defined as the probability that a call is blocked after waiting a specific length of time in the queue. To find the GOS, it is first necessary to find the likelihood that a call is initially denied access to the system. The likelihood of a call not having immediate access to a channel is determined by the Erlang C formula derived in Appendix A

$$Pr[\text{delay} > 0] = \frac{A^C}{A^C + C! \left(1 - \frac{A}{C}\right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \quad (2.17)$$

If no channels are immediately available the call is delayed, and the probability that the delayed call is forced to wait more than t seconds is given by the probability that a call is delayed, multiplied by the conditional probability that the delay is greater than t seconds. The GOS of a trunked system where blocked calls are delayed is hence given by

$$\begin{aligned} Pr[\text{delay} > t] &= Pr[\text{delay} > 0] Pr[\text{delay} > t | \text{delay} > 0] \\ &= Pr[\text{delay} > 0] \exp(-(C-A)t/H) \end{aligned} \quad (2.18)$$

The average delay D for all calls in a queued system is given by

$$D = Pr[\text{delay} > 0] \frac{\bar{H}}{C-A} \quad (2.19)$$

Where the average delay for those calls which are queued is given by $H/(C-A)$.

The Erlang B and Erlang C formulas are plotted in graphical form in Figure 2.6 and Figure 2.7. These graphs are useful for determining GOS in rapid fashion, although computer simulations are often used to determine transient behaviors experienced by particular users in a mobile system.

To use Figure 2.6 and Figure 2.7, locate the number of channels on the top portion of the graph. Locate the traffic intensity of the system on the bottom portion of the graph. The blocking probability $Pr[\text{blocking}]$ is shown on the abscissa of Figure 2.6, and $Pr[\text{delay} > 0]$ is shown on the abscissa of Figure 2.7. With two of the parameters specified it is easy to find the third parameter.

Example 2.4

How many users can be supported for 0.5% blocking probability for the following number of trunked channels in a blocked calls cleared system? (a) 1, (b) 5, (c) 10, (d) 20, (e) 100. Assume each user generates 0.1 Erlangs of traffic.

Solution to Example 2.4

From Table 2.4 we can find the total capacity in Erlangs for the 0.5% GOS for different numbers of channels. By using the relation $A = UA_u$, we can obtain the total number of users that can be supported in the system.

(a) Given $C = 1$, $A_u = 0.1$, $GOS = 0.005$

From Figure 2.6, we obtain $A = 0.005$.

Therefore, total number of users, $U = A/A_u = 0.005/0.1 = 0.05$ users.

But, actually one user could be supported on one channel. So, $U = 1$.

(b) Given $C = 5$, $A_u = 0.1$, $GOS = 0.005$

From Figure 2.6, we obtain $A = 1.13$.

Therefore, total number of users, $U = A/A_u = 1.13/0.1 \approx 11$ users.

(c) Given $C = 10$, $A_u = 0.1$, $GOS = 0.005$

From Figure 2.6, we obtain $A = 3.96$.

Therefore, total number of users, $U = A/A_u = 3.96/0.1 \approx 39$ users.

(d) Given $C = 20$, $A_u = 0.1$, $GOS = 0.005$

From Figure 2.6, we obtain $A = 11.10$.

Therefore, total number of users, $U = A/A_u = 11.1/0.1 = 110$ users.

(e) Given $C = 100$, $A_u = 0.1$, $GOS = 0.005$

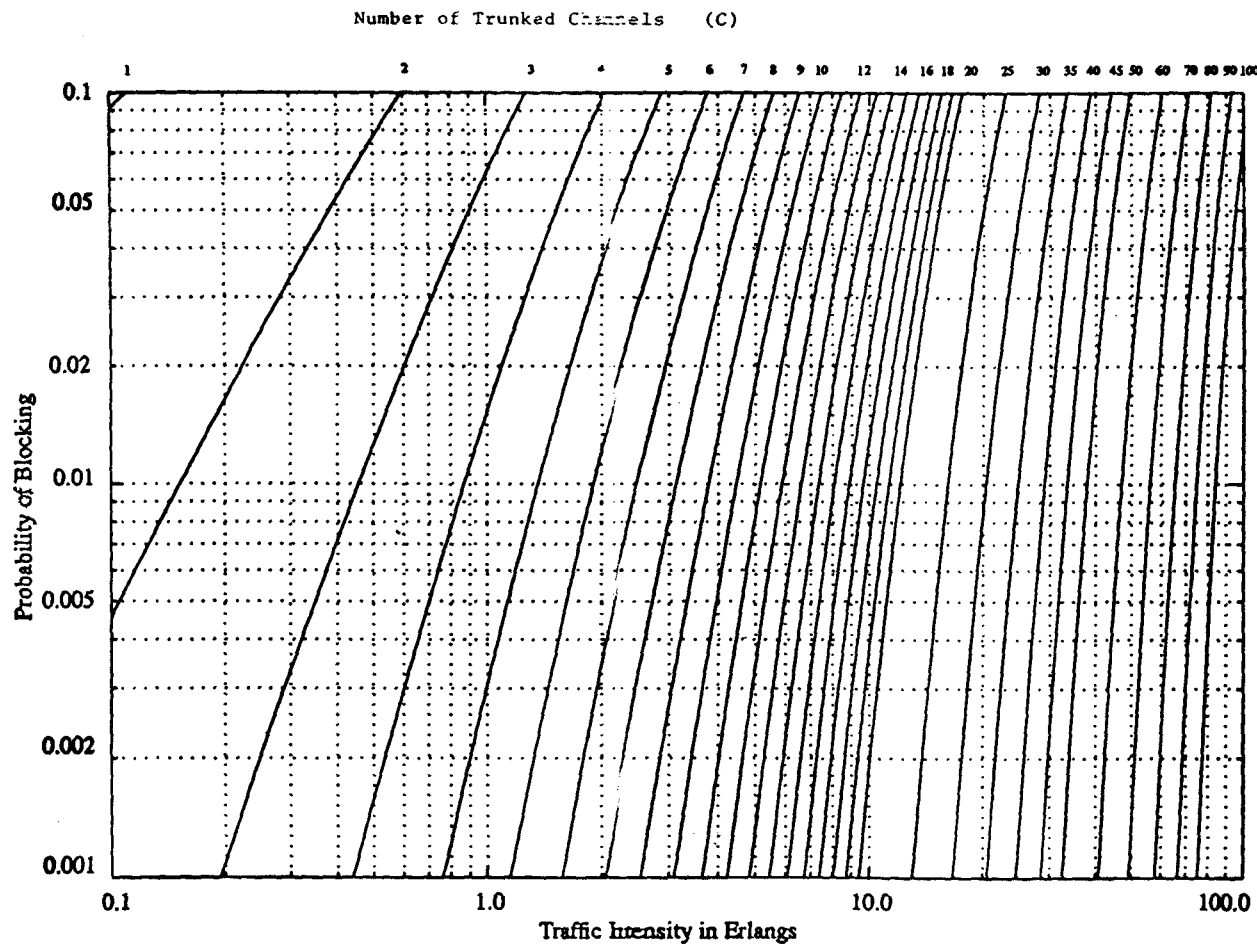


Figure 2.6
The Erlang B chart showing the probability of blocking as functions of the number of channels and traffic intensity in Erlangs.