Trunking and Grade of Service

Trunking refers to the sharing of a number of channels ("trunks" in telephone network terminology) among a larger number of users. Users seize a trunk at the start of a call and release it when the call is complete.

There is a possibility that all trunks will be in use when a user wants to place a call. In this case the call is "blocked".

The probability of blocking or the mean waiting time will be a function of the number of channels available, the call volume and mean call duration.

Trunking theory is part of queueing theory. Queueing theory studies the statistics of waiting and service times, queue lengths, blocking probabilities, etc.

The traffic intensity is measured in Erlangs which is the average utilization of a channel. For example a channel which is in use 90% of the time is carrying 0.9 Erlangs.

Grade of Service (GoS) is the probability that a call will be blocked. It is usually specified for the average offered load during the busiest hour of the week.

The offered load per user is specified in Erlangs. It is specified as:

$$A_u = \lambda H$$

where λ is the service request rate in calls/hour (or per minute) and H is the mean holding time (call duration) in hours (or minutes). The product is in units of Erlangs. If there are U users in a call the total offered load is $A = UA_u$. If there are C channels the offered load per channel is $A_C = A/C$.

Cellular systems are provisioned (by cell sizing and channel allocations) to provide a GoS of a few percent (2%). This means during the busiest hour about 2% of calls will not go through because all channels are busy.

The simplest model for estimating blocking probability is to assume:

- an infinite number of users and a finite number of channels (or at least, much larger than the number of channels)
- users that are blocked give up, they are not queued and the call blocking does not affect the time of the next call ("block calls cleared")
- there is a Poission-distributed call arrival rate λ (equivalently, the inter-arrival times are exponentially distributed since the probability of a user making a call in any given time interval is constant)
- the call holding time is exponentially distributed

Under these conditions the blocking probability is the GoS and is given by the Erlang-B formula:

$$P_b = \frac{\frac{A^C}{C!}}{\sum_{k=0}^{C} \frac{A^k}{k!}}$$

where C is the number of channels, and A is the offered load.

The following graph shows the blocking probability using the Erlang-B formulas for various numbers of channels C and offered load A.

Exercise 1: A cell has 20 channels available for 1200 users. Each users attempts to make one call every two hours and the average call time is 1.2 minutes. What is the probability that a call will be blocked? What if there were 2 channels for 100 users?

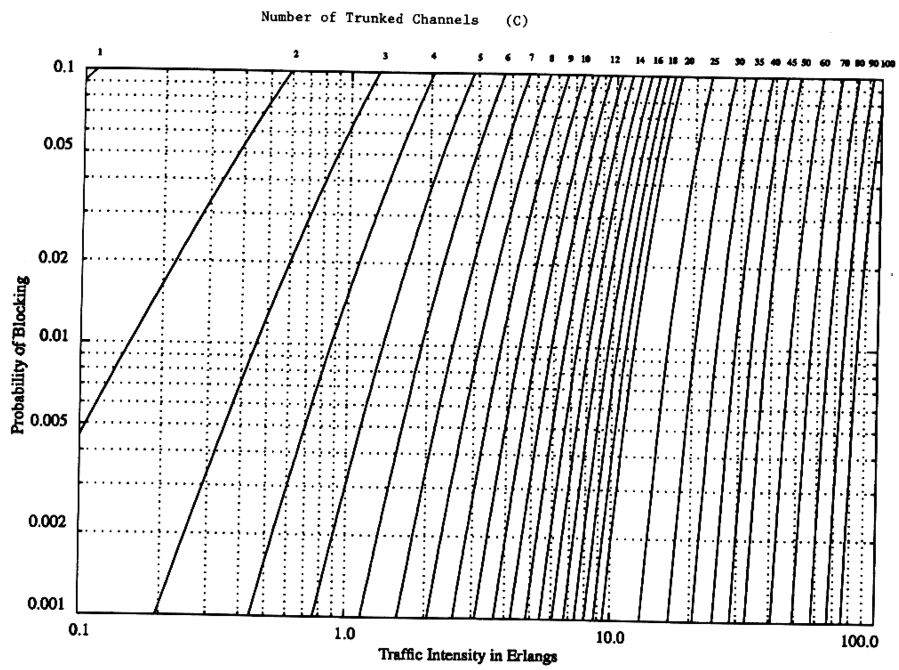

Figure 3.6 The Erlang B chart showing the probability of blocking as functions of the number of channels and traffic intensity in Erlangs.