Teletraffic Engineering

January 2016

Traffic Engineering

- Traffic engineering uses *statistical techniques* such as queuing theory to *predict and engineer* the behavior of telecommunications networks such as telephone networks or the Internet.
  - One of the important steps of teletraffic engineering determines *number of channels* required on a route or a connection between two MSCs [1].
  - Another important steps of teletraffic engineering is to ensure the desired

[1] Mobile switching centre server which is a 3G term

Busy hour

- Busy hour - uninterrupted period 60 min during which the traffic volume is highest
- Used for traffic dimensioning
- Can be: fixed/mobile
- In heterogeneous networks busy hours for different traffic types may not coincide
- In cellular network, busy hour occurs at different time for different cells

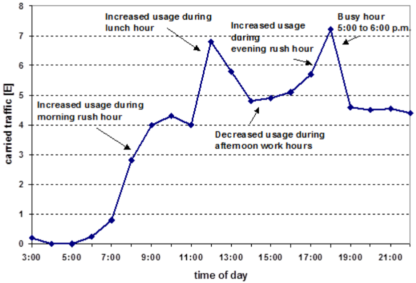

Figure: Typical daily traffic usage in a cellular system.

Service Level

Service Level can be divided into *two main areas*:

1. **Dial tone delay:** The maximum waiting time to hear a dial tone after removing the hand-set from the hook.
2. **Service blocking probability:**
   - the probability that service delay will exceed some specific value or
   - The probability that the call will be denied or blocked

Service or call blocking probability is known as the **grade of service (GoS)**

$$GoS = \frac{Block\ \ Call}{Total\ \ Call} \tag{1}$$

Serviced Calls = 380, blocked calls = 10, GoS?

$$GoS = \frac{Block\ \ Call}{Serviced\ \ calls + Blocked\ \ calls} = \frac{10}{380 + 10} = \frac{1}{39} \tag{2}$$

Teletraffic Engineering - Definitions

- *Set-up Time:* The time required to allocate a radio channel to a requesting user.
- *Blocked Call:* Call which cannot be completed at time of request, dues to congestion. Also referred to as a *lost call*.
- *Holding Time:* Average duration of typical call (Denoted by $H = 1/\mu$).
- *Traffic Intensity:* Measure the 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 a single or mutiple channels (Denoted by A or ρ).
- *Request Rate:* The average number of call requests per unit time (Denoted by λ).

$$\text{A} = \lambda * H = \frac{\lambda}{\mu} = \rho$$

Teletraffic Engineering - Definitions

- *Grade of Service (GOS):* 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).
- *Load:* Traffic intensity across the system, measured in Erlangs.
- *Erlangs:* - describe traffic intensity in terms of the number of hours of resource time required per hour of elapsed time
- *Centum Call Seconds (CCS):* - measures the exact same traffic intensity as the Erlangs but expresses it as the number of 100 second holding times required per hour. Traffic registers sample stations every 100 seconds per hour to check for busies. Since there are 36 sets of hundred seconds in an hour (CCS = 36 x Erlangs)

Teletraffic Engineering - Erlangs

Figure: Named after Danish mathematician A. K. Erlang (1878-1929)

Defined as one circuit occupied for one hour. 1 Erlang= 1 Callhour / hour

Busy hour traffic

$$\text{Erlangs} = (\text{Calls/busy hour})*(\text{ call holding time})$$

Example A group of 20 subscribers generate 50 calls with an average holding time of 3 minutes, what is the average traffic per subscriber?

- Total Traffic = (50 calls)*(3min)/(1 hour)=50*3/60 = 2.5 Erlangs
- = 2.5 / 20 or 0.125 Erlangs per subscriber.

Teletraffic Engineering - Birth-Death

A *Birth-Death* is a special type of discrete-time or continuous time Markov Chain with the restriction that at each step, the state transitions, if any, can occur only between neighbouring states.

The underlying Markov process representing the number of customers in such systems is know as a *birth-and-death* process. The birth-death terminology is used to represent increases and decreases in the population size. The corresponding events in queueing systems are arrivals and departures.

Again using the birth (arrival)-death (departure) terminology, when the population size is n, let λ_n and μ_n be the infinitesimal transition rates (generators) of birth and death, respectively.

When the population is the number of customers in the system, λ_n and μ_n indicate that the arrival and service rates depend on the number in the system.

If the process is a Birth-death process and if the current state X_n is i, then the above condition implies that the next state X_{n+1} can only be i-1, i or i+1.

Based on the properties of the Poisson process, i.e., when arrivals are in a Poisson process and service times are exponential, we can make the following probability statements for a transition during $(t, t + \Delta t]$:

birth(n $\geq$ 0):

$$\begin{aligned} P(\text{one birth}) &= \lambda_n \Delta t + o(\Delta t) \\ P(\text{no birth}) &= 1 \text{ - } (\lambda_n \Delta t + o(\Delta t)) \\ P(\text{more than one birth}) &= o(\Delta t) \end{aligned}$$

death(n $>$ 0):

$$\begin{aligned} P(\text{one death}) &= \mu_n \Delta t + o(\Delta t) \\ P(\text{no death}) &= 1 \text{ - } (\mu_n \Delta t + o(\Delta t)) \\ P(\text{more than one death}) &= o(\Delta t) \end{aligned}$$

where $o(\Delta t)$ is such that $\frac{o(\Delta t)}{\Delta t} n \to 0$ as $\Delta t \to 0$. Since a continouse-time process is being considerd, we need to focus on changes in the process over time interval $o(\Delta t)$ as $o(\Delta t) \to 0$. Let λ_k be the birth rate in state k[2] Similarly, let μ_k be the death rate in state k.

[2]Note that is these statements the $o(\Delta t)$ terms do not specify actual values. In each of the two cases, the $o(\Delta t)$ terms sum to 0 so that the total probability of the three events is equal 1.

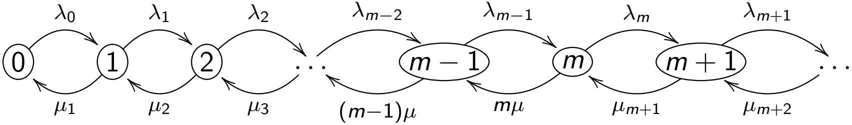

Figure: State Transition Diagram (Birth-Death)

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} -\lambda_0 & \lambda_0 & \cdots & \cdots & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \cdots & \cdots \\ \vdots & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

Notation for Queues

The notation is A/B/C/D/E:

A		This symbolically represents the nature of the arrival process to the queue. Special letters are used to symbolise the nature of the inter-arrival time distribution. Some of the important ones are
	M	Exponentially distributed inter-arrival times (Poisson Process[3])
	D	Deterministic (fixed) inter-arrival times
	E_k	Erlang distribution of order k for inter-arrival times
	H_k	Hyper-exponential distribution of order k for inter-arrival times
	G	General (any!) distribution for the inter-arrival times
B		This symbolically represents the nature of the service time distribution for the customers getting served in the queue. The same letters as the ones above are used to describe the nature of the service time distribution
C		Number of servers in the queue

[3] Poisson process is a collection $(N(t) : t \geq 0)$ of random variables, where $N(t)$ is the number of events that have occurred up to time t (starting from time 0).

Notation for Queues

The notation is A/B/C/D/E:

D Maximum Number of jobs/customers that can be there in the system. This includes both the ones currently being served and the ones waiting for service. Note that the default is infinity (∞) which is assume when this is omitted.

E Queueing Discipline such as -
FCFS First Come First Served
LCFS Last Come First Served
SIRO Service In Random Order
this may also be omitted if the queue is FCFS in nature (default)

Notation for Queues

Examples:

M/M/1 or $M/M/1/\infty$	Poisson Arrivals, Exponential Service Time Distribution [4], Single Server, Infinite Number of Waiting Positions
$M/E_2/2/K$	Poisson Arrivals, Erlangian of order-2 Service Time Distribution, Two Servers, Maximum Number K in system (waiting and in service)
G/M/2	Generalised Arrivals, Exponential Service Time Distribution, 2 Servers, Infinite Number of Waiting Positions

[4] It describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. In other words it is a continuous probability distribution whose density function is given by $f(x) = ae^{-ax}$, where $a > 0$ for $x > 0$, and $f(x) = 0$ for $x \leq 0$; the mean and standard deviation are both $1/a$.

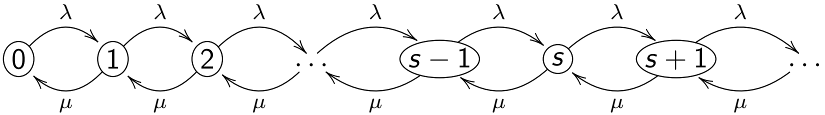

Figure: State Transition Diagram ($M/M/1$)

$$
\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\
\begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \cdots \\
\mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\
0 & \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\
0 & 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\end{array}
$$

Teletraffic Engineering - $M/M/\infty$

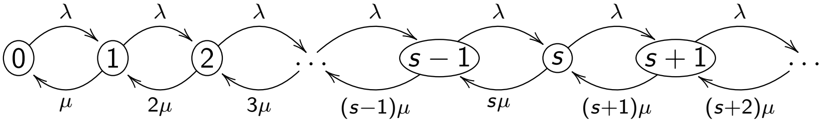

Figure: State Transition Diagram ($M/M/\infty$)

$$
\begin{array}{c}
\\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots
\end{array}
\begin{array}{c}
\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\
\begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \cdots \\
\mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\
0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 & \cdots \\
0 & 0 & 3\mu & -(\lambda+3\mu) & \lambda & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\end{array}
$$

Teletraffic Engineering - $M/M/s$

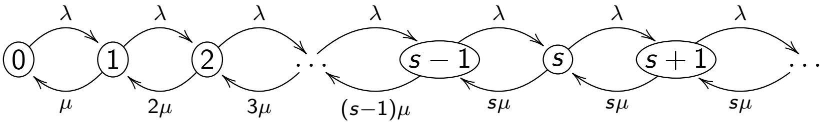

Figure: State Transition Diagram ($M/M/s$)

$$\begin{array}{c} 0 \\ 1 \\ \vdots \\ s \\ s+1 \\ \vdots \end{array}
\begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \cdots & \cdots & \cdots \\
\mu & -(\lambda+\mu) & \lambda & 0 & \ddots & \ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots & s\mu & -(\lambda+s\mu) & \lambda & \ddots & \ddots \\
\ddots & \ddots & \ddots & \ddots & s\mu & -(\lambda+s\mu) & \lambda & \ddots \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots
\end{pmatrix}$$

Teletraffic Engineering - $M/M/1/K$

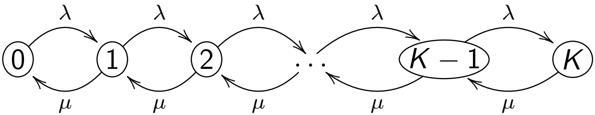

Figure: State Transition Diagram ($M/M/1/K$)

$$
\begin{matrix} 0 \\ 1 \\ \vdots \\ K \end{matrix}
\begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \cdots & \cdots & 0 \\
\mu & -(\lambda+\mu) & \lambda & 0 & \ddots & \ddots & \ddots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots & \mu & -(\lambda+\mu) & \lambda
\end{pmatrix}
$$

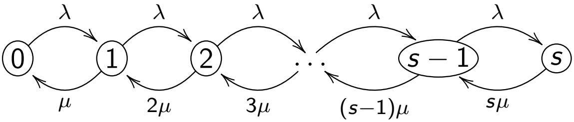

Figure: State Transition Diagram ($M/M/s/s$)

$$
\begin{array}{c}
0 \\ 1 \\ \vdots \\ s
\end{array}
\begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \cdots & \cdots & 0 \\
\mu & -(\lambda+\mu) & \lambda & 0 & \ddots & \ddots & \ddots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots & s\mu & -(\lambda+s\mu) & \lambda
\end{pmatrix}
$$

Erlang–B Formula[5]

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

Call blocking probability

C = Number of trunked channels

$A =$ traffic intensity or load [Erlangs]

$$= \frac{\lambda}{\mu}$$

$\lambda =$ Average # call attempts/requests per unit time

$$\frac{1}{\mu} = H = \text{Average call length}$$

[5] Also known as Erlang-1, Erlang-loss

Erlang-B Example

A cell site has 5 FDMA radios. The average call origination rate is 60 calls per hour. If the call holding times are distributed exponentially with an average of 90 sec, calculate the GOS

Average birth rate $\lambda = 60$ calls/hour

Average death rate $\mu = \frac{1}{H} = \frac{1}{90/3600} = 40$ calls/hour

Traffic Intensity $\text{A} = \frac{\lambda}{\mu} = \frac{60}{40} = 1.5$ Erlangs

GoS

$$B\left[A = 1.5, C = 5\right] = \frac{A^C/C!}{\sum_{i=0}^{C} \frac{A^i}{i!}}$$

$$= \frac{1.5^5/5!}{\frac{1.5^0}{0!} + \frac{1.5^1}{1!} + \frac{1.5^2}{2!} + \frac{1.5^3}{3!} + \frac{1.5^4}{4!} + \frac{1.5^5}{5!}} \approx 0.0142$$

Cell Capacity

- Average number of MSs requesting service (Average arrival rate): λ
- Average length of time MS requires service (Average holding time): T
- *Offered* load: A $= \lambda T$ where a is in Erlangs

e.g., in a cell with 100 MSs, on an average 30 requests are generated during an hour, with average holding time T=360 seconds

Then, arrival rate λ=30/3600 requests/sec

A completely occupied channel (1 call-hour per hour) is defined as a load of one Erlang, i.e.,

$$A = \frac{30\ calls}{3600\ sec} \cdot \frac{360\ sec}{call} = 3\ Erlangs$$

Capacity of a Cell

- This is Erlang B formula B(C, A)
- In the previous example, if C = 2 and A = 3, the blocking probability B(2, 3) is

$$B\left[2,3\right] = \frac{A^C/C!}{\sum_{i=0}^{C}\frac{A^i}{i!}} = \frac{3^2/2!}{\sum_{i=0}^{2}\frac{3^i}{i!}} = 0.529$$

- So, the number of calls blocked $30 \cdot 0.529 = 15.87$