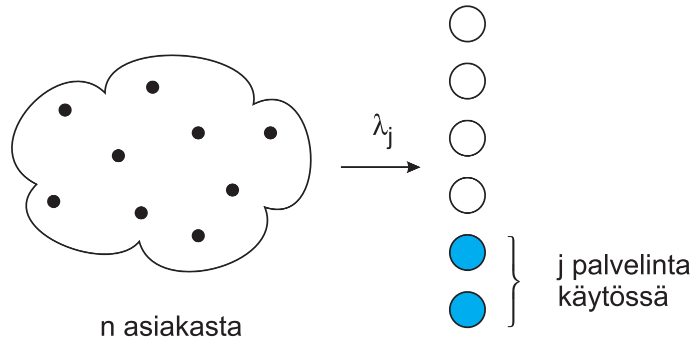
Finite source population: the $M/M/s/s/n$ system

Consider the loss system (no waiting places) in the case where the arrivals originate from a finite population of sources: the total number of customers is n.



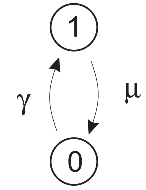
The customer model

- To be specific, think of the customers being telephone users.
- Assume that the time to the next call attempt by the customer, so called thinking time (idle time) of the customer obeys the distribution Exp(γ).
- Blocked calls are lost
  - does not lead to reattempts
  - starts a new thinking time: again, the time to the next attempt $\sim$ Exp(γ)
  - the holding time $X \sim$ Exp(μ)

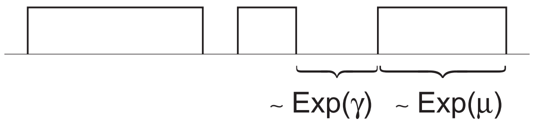
Denote $\boxed{\hat{a} = \gamma/\mu}$

Behaviour of a single source

<u>The case $s = \infty$</u> $(s \geq n)$ Each customer has his own server.



- No blocking; all call attempts are accepted.
- The user behaves according to the two-state (0,1)-model of the figure. Alternating active (1) period $\sim$Exp(μ) and thinking (0) period $\sim$Exp(γ).



$$p_0 = \frac{1/\gamma}{1/\gamma + 1/\mu} = \frac{1}{1+\hat{a}}, \quad p_1 = \frac{1/\mu}{1/\gamma + 1/\mu} = \frac{\hat{a}}{1+\hat{a}}$$

- The mean duration of a full cycle (thinking period + active period) is $1/\gamma + 1/\mu$.
- The arrival rate of calls (one per cycle) is

$$\frac{1}{1/\gamma + 1/\mu} = \frac{1/\gamma}{1/\gamma + 1/\mu}\gamma = p_0\gamma$$

the arrival rate in the thinking state is γ
the probability of the thinking state is p_0

- Offered load = carried load (per source)

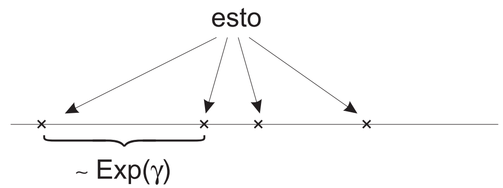
$$a^{\infty} = p_1 = \frac{\hat{a}}{1+\hat{a}}$$

so called intended load per source (no blocking)

The behaviour of a single source (continued)

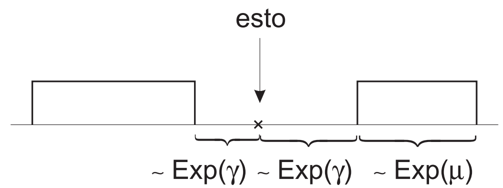
The case $s = 0$: No servers.

- All call attempts are blocked.
- Call attempts are generated at the rate γ, which is greater than the rate $(p_0\gamma)$ in the case $s = \infty$, since now the customer spends all the time in the thinking state.



The case $0 \le s \le n$: This the case of interest.

- The state of the customer alternates between an active state $\sim$Exp(μ) and a thinking state which consists of one or more periods obeying the distribution Exp(γ).



The analysis of the system as a birth-death process

Take the number on calls in progress, N_t, as the state variable.

N_t indeed constitutes a Markov process of the birth-death type.

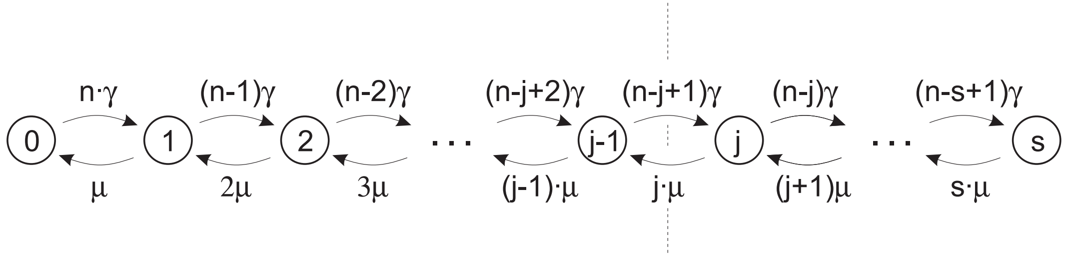
The state variable N_t determines (irrespective of the past) the probabilities of arrival of a new call or departure of an on-going call per time unit (memoryless exponential distributions!).

In the state $N_t = j$

- $n - j$ source free in the thinking state
  - the time to the generation of the next call attempt $\sim \text{Exp}((n-j)\gamma)$
  - the probability per time unit for the transition to a higher occupancy state $\lambda_j = (n-j)\gamma$ (state dependent arrival rate)
  - in the state s all servers are occupied, whence $\lambda_s = 0$
- j calls in progress
  - time to the next end of a call $\sim \text{Exp}(j\mu)$
  - $\mu_j = j\mu$ (state dependent departure rate)

The equilibrium probabilities

The birth-death process is described by the following state transition diagram



By equating the probability flows across the cut, one gets the recursion

$$(n-j+1)\,\gamma\,\pi_{j-1} = j\mu\,\pi_j, \qquad j = 1, 2, \ldots, s$$

by which all probabilities can be reduced to π_0,

$$\pi_j = \frac{n-j+1}{j} \cdot \frac{n-j+2}{j-1} \cdots \frac{n-1}{2} \cdot \frac{n}{1} \cdot \left(\frac{\gamma}{\mu}\right)^j \pi_0 = \binom{n}{j} \hat{a}^j\, \pi_0$$

By applying the normalization condition $\sum_{j=0}^{s} \pi_j = 1$

$$\boxed{\pi_j[n] = \frac{\binom{n}{j} \hat{a}^j}{\sum_{k=0}^{s} \binom{n}{k} \hat{a}^k}} \qquad j = 0, 1, \ldots, s$$

where the number of sources, n, is explicitly shown, $\pi_j[n]$

Equilibrium probabilities (continued)

Denote $p = p_1 = a^\infty$, the on-probability of a source in the non-blocking case.

$$p = \frac{\hat{a}}{1+\hat{a}} \qquad \Rightarrow \qquad \hat{a} = \frac{p}{1-p}$$

In terms of this the expression for the state probabilities can be written as

$$\boxed{\pi_j[n] = \frac{\binom{n}{j} p^j (1-p)^{n-j}}{\sum_{k=0}^{s} \binom{n}{k} p^k (1-p)^{n-k}}} \qquad \begin{array}{l} j = 0, 1, \ldots, s \\ \\ \text{(0 otherwise)} \end{array}$$

- Truncated binomial distribution (when $s < n$)
- Binomial distribution, when $s \geq n$
  (non-blocking case, each source is in the on-state with the probability p independent of the others)

Insensitivity: In the same way as in the ordinary loss system, the result is insensitive to the form of the holding time distribution (though the derivation above was explicitly based on the assumption of exponential holding time distribution).

Time and call blocking probabilities

Time blocking: the proportion of time the system spends in the state s; the equilibrium probability π_s of the state s

$$\boxed{E_n(s,\hat{a}) = \frac{\binom{n}{s}\hat{a}^s}{\sum_{k=0}^{s}\binom{n}{k}\hat{a}^k}}$$

Call blocking: the probability that an arriving call is blocked

- Because of the state dependent arrival rate the arrival process is not Poissonian.
- Call blocking $\neq$ time blocking
- Generally $\pi_j^*[n] \neq \pi_j[n]$, where

$$\begin{cases} \pi_j^*[n] = \mathrm{P}\{\text{upon an arrival the system is in the state } j\} \\ \pi_j[n] = \mathrm{P}\{\text{the system is in the state } j \text{ at a random instant of time}\} \end{cases}$$

The state probabilities seen by an arriving call

$$\boxed{\pi_j^*[n] = \frac{\lambda_j \pi_j[n]}{\sum_{k=0}^{s} \lambda_k \pi_k[n]}}$$

Proof: Consider a long period of time T.

- On the average the system spends in state j the time $\pi_j[n]T$.
- During this time there are on the average $\lambda_j \pi_j[n]T$ call arrivals (which find the system in the state j).
- The total number of calls arriving in time T is on the average $T \sum_{k=0}^{s} \lambda_k \pi_k[n]$.
- The proportion of calls which find the system in the state j is as given by the expression.

By inserting in the above expression $\lambda_j = (n-j)\gamma$ one finds (proof left as an exercise) that

$$\boxed{\pi_j^*[n] = \pi_j[n-1]}$$

The state distribution seen by an arriving customer is the same as the equilibrium distribution in a system with one less customer. It is as if the arriving customer were an "outside observer".

Call blocking:

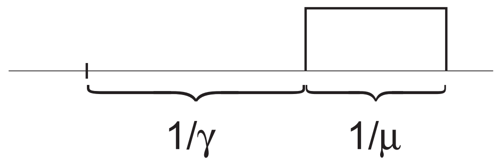
(Engset's formula)

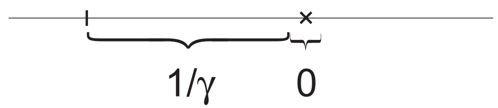
$$\boxed{B_n(s,\hat{a}) = \pi_s^*[n] = \pi_s[n-1] = \frac{\binom{n-1}{s}\hat{a}^s}{\sum_{k=0}^{s}\binom{n-1}{k}\hat{a}^k}}$$

The realized offered and carried loads

Denote the call blocking probability for short by B.
Consider one cycle which consists of

- Thinking period (mean duration $1/\gamma$)
- Service period which
  - with the probability $1-B$ is a real active period (mean duration $1/\mu$)
  - with the probability B is shrunk to blocking event of duration 0





The mean duration of a cycle is $\quad 1/\gamma + (1-B)\cdot 1/\mu + B\cdot 0 = 1/\gamma + (1-B)\cdot 1/\mu.$

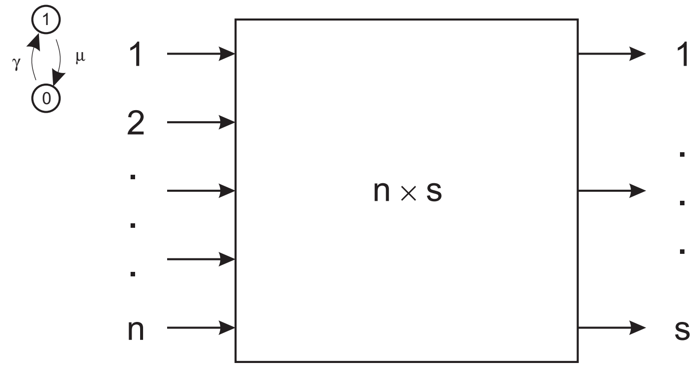
The probability that the source is in the thinking phase is $\quad \dfrac{1/\gamma}{1/\gamma + (1-B)\cdot 1/\mu} = \dfrac{1}{1+(1-B)\hat{a}}$

Each of the n sources when being in the thinking phase generates new call attempts at the rate γ. The offered traffic intensity a is the arrival rate times the mean holding time of a call $1/\mu$.

$$\boxed{a = \frac{\hat{a}}{1+(1-B)\hat{a}}\, n}$$

The carried traffic intensity $\quad \boxed{a_c = (1-B)a = \dfrac{(1-B)\hat{a}}{1+(1-B)\hat{a}}\, n} \quad = \sum_{k=0}^{s} k\,\pi_k$

Example. A finite population system: concentrator



- n users, each of which is connected by an own line to the input side of the concentrator.
- $s < n$ output lines; the traffic is concentrated on a smaller number of trunks, because it is unlikely that all the users are active at the same time.
- Simultaneous use, however, is possible; the degree of concentration has to be dimensioned such that the probability of blocking is small enough.
- The blocking can be calculated with the Engset formula for the call blocking.