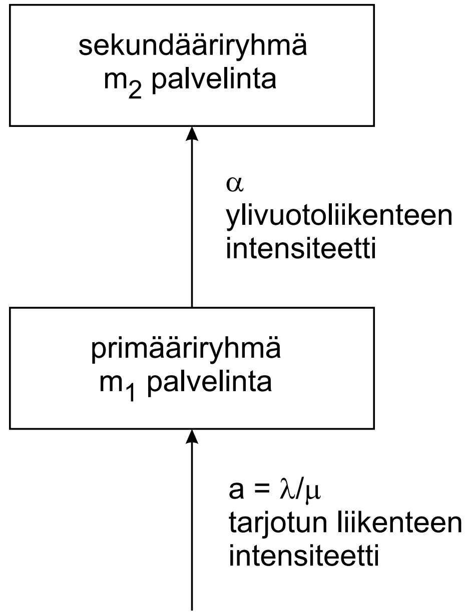
OVERFLOW TRAFFIC

Overflow traffic in a loss system

Consider a circuit switched network which works as a loss system (blocked calls are cleared).

- System is offered Poissonian traffic with intensity λ.
- The holding times are exponentially distributed with parameter μ.
- The offered traffic intensity $a = \lambda/\mu$.

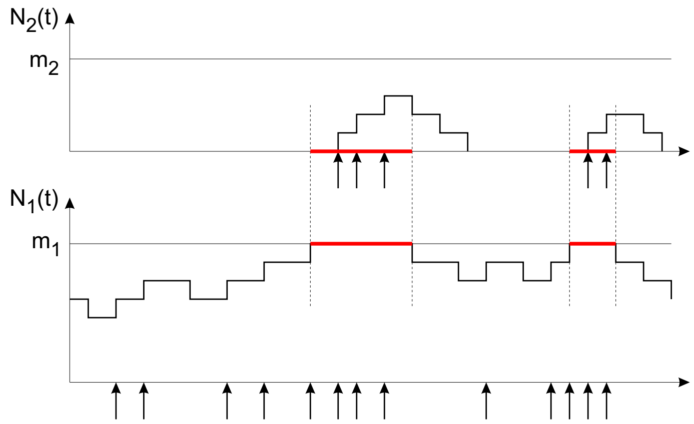


The system consists of two parts:

- primary system: m_1 servers
- secondary system: m_2 servers

- The arriving traffic is first offered to the primary system.
- If all servers of the primary group are occupied the call is directed to the secondary group,
  – the traffic directed from the primary group to the secondary group is called overflow traffic; denote its intensity by α.
- If all servers of the secondary group are also occupied, the arriving call is blocked.

Traffic process of the overflow traffic



The primary and secondary groups together constitute a m_1+m_2 server loss system (Erlang's system; $M/M/m/m$ queue, where $m = m_1 + m_2$).

The primary system alone forms an m_1 server loss system ($m = m_1$).

The division of the capacity m_1+m_2 of the system into two parts does not affect the overall behaviour of the system (the blocking probability is the same as in an m_1+m_2 server system).

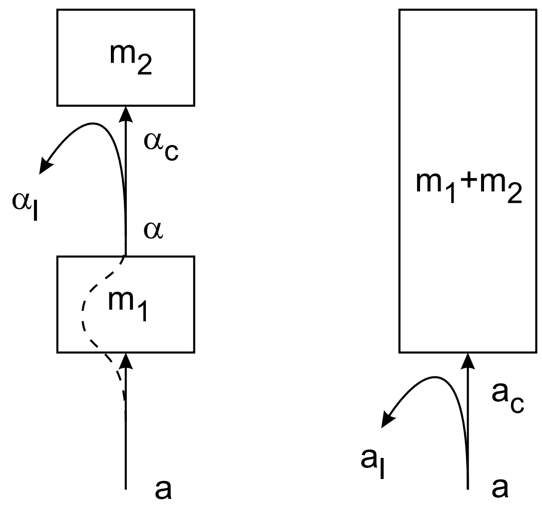
One can think the trunks of the system to be numbered. An arriving call is carried by the free trunk with the lowest number.

$$\begin{cases} N_1 & \sim \text{ truncated } (m_1) \text{ Poisson distribution} \\ N = N_1 + N_2 & \sim \text{ truncated } (m_1+m_2) \text{ Poisson distr.} \\ N_2 = N - N_1 & \quad \text{occupancy of the secondary group} \end{cases}$$

Note. Though both N and N_1 are separately insensitive (independent of the holding time distribution), the distribution of N_2 is not insensitive.

Overflow traffic is generated only when the primary group is in blocking state. The blocking periods of the primary group form time windows, during which the arriving traffic is directed to the secondary group. The arrival process of the overflow traffic is so called interrupted Poisson process (IPP).

Blocking of the overflow traffic



Intensity of the overflow traffic α
(traffic "blocked" in the primary group)

$$\alpha = a \cdot E(m_1, a)$$

where $E(m, a)$ is the Erlang B-formula.

Ultimately blocked traffic $\alpha_\ell = a_\ell$
(traffic blocked in the secondary group)

$$a_\ell = a \cdot E(m_1 + m_2, a)$$

Call blocking of the overflow traffic

$$B_2 = \frac{aE(m_1 + m_2, a)}{aE(m_1, a)}$$

$$\boxed{B_2 = \frac{E(m_1 + m_2, a)}{E(m_1, a)}}$$

One can show that $B_2 > E(m_2, \alpha)$

The blocking experienced by the overflow traffic is greater than the blocking that would be experienced by Poisson traffic with the same intensity in a secondary group of m_2 trunks.

This is due to the fact that the IPP process is more bursty than the ordinary Poisson process.

Blocking of the overflow traffic (continued)

Example

Assume that the intensity of the offered traffic is $a = 5$ erl.

The sizes of the primary and secondary groups: $m_1 = 5$ and $m_2 = 1$

$$\begin{cases} \text{overflow traffic} & \alpha = 5 \cdot E(5,5) = 5 \cdot 0.285 = 1.42 \\ \text{finally blocked traffic} & a_\ell = 5 \cdot E(6,5) = 5 \cdot 0.192 = 0.96 \end{cases}$$

$$\begin{cases} \text{blocking of overflow traffic} & B_2 = \dfrac{a_\ell}{\alpha} = \dfrac{E(6,5)}{E(5,5)} = 0.67 \\ \text{blocking if the traffic were Poissonian} & E(1, 1.42) = 0.59 \end{cases}$$

Peakedness of the overflow traffic

Sometimes it is useful to characterize a traffic process by telling what would be the occupancy distribution if the traffic were offered to a trunk group of infinite capacity.

To characterize the overflow traffic in this way, assume now the secondary group is infinite, $m_2 = \infty$.

Then all the overflow traffic will be carried in the secondary group and it holds

$$\boxed{\mathrm{E}[N_2] = aE(m_1, a) = \alpha}$$

Also the variance of N_2 can be calculated in this case. The derivation is rather involved. The result is known as the Riordan formula:

$$\boxed{\mathrm{V}[N_2] = \alpha\left(1 - \alpha + \frac{a}{m_1 + 1 - a + \alpha}\right)}$$

The variance to mean ratio of the occupancy is called the peakedness factor. (In the case of a Poisson arrival process the occupancy distribution is Poisson(a) distribution with the peakedness factor 1.)

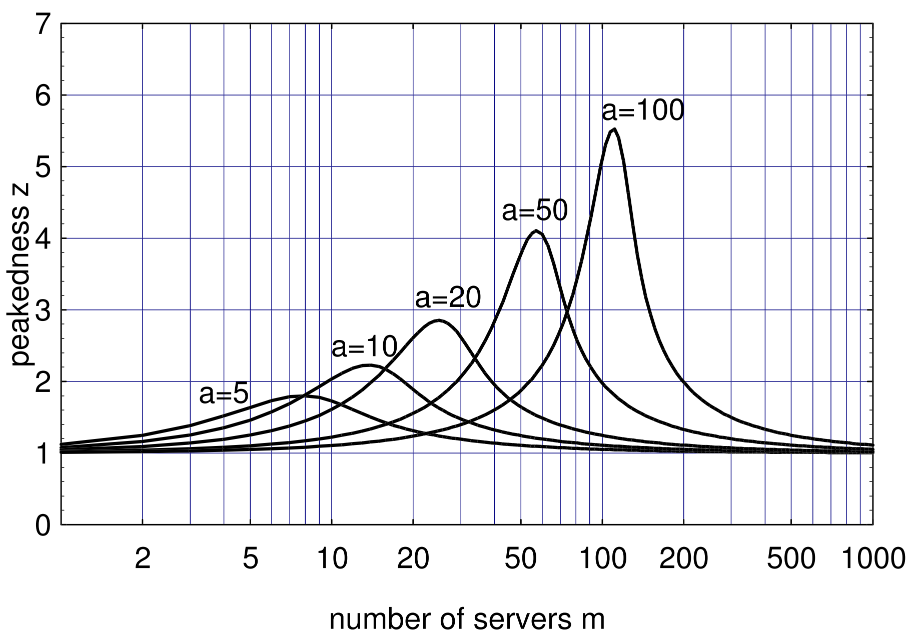
$$\boxed{z = \frac{\mathrm{V}[N_2]}{\mathrm{E}[N_2]} = 1 - \alpha + \frac{a}{m_1 + 1 - a + \alpha}} \qquad \text{where } \alpha = aE(m_1, a)$$

Peakedness of the overflow traffic (continued)

The peakedness factor is a function of m_1 and a, $z = z(m_1, a)$.

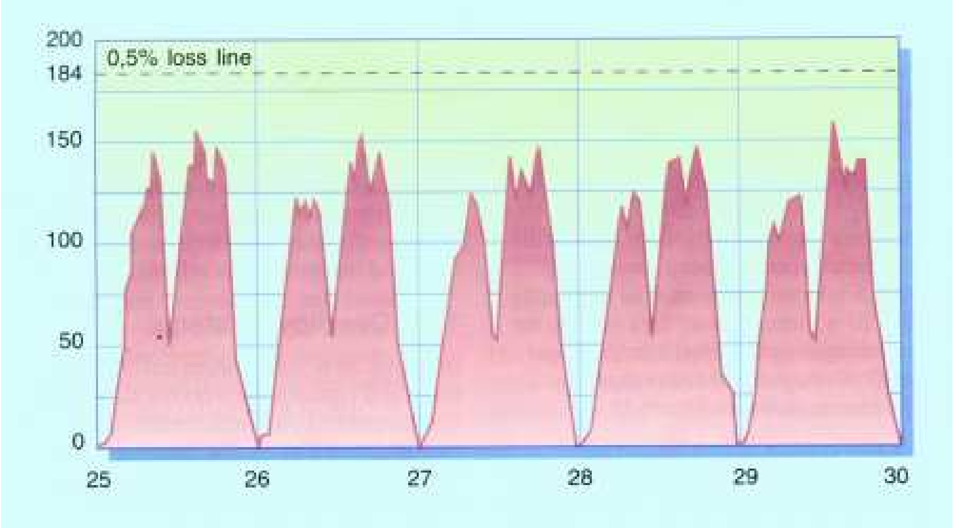
When a is held fixed and m_1 is increased

- first, for small m_1, $z \approx 1$ (all the traffic flows over)
- then z attains a maximum (when $m_1 \approx a$)
- finally, for very large m_1, $z \to 1$ (the overflow events are rare singular events)



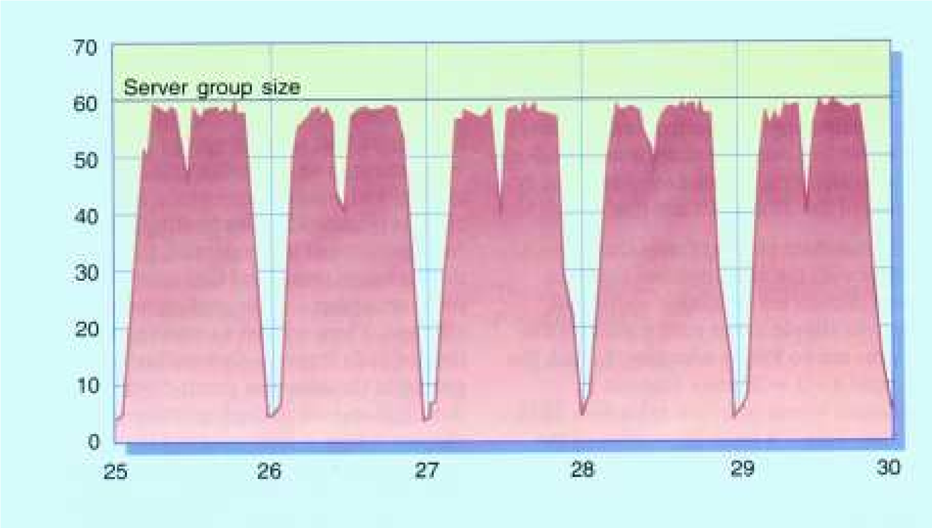
An example of overflow traffic

The figure on the right shows the offered traffic in a period of five days [a]



The figures below show the same traffic in a primary group of 60 trunks and the overflow traffic. The traffic in the primary group is smoother (truncated) than the offered traffic, whereas the overflow traffic is more peaky than the offered traffic.

[a] From: A. Myskja, Telektronikk 2/3 (1995).



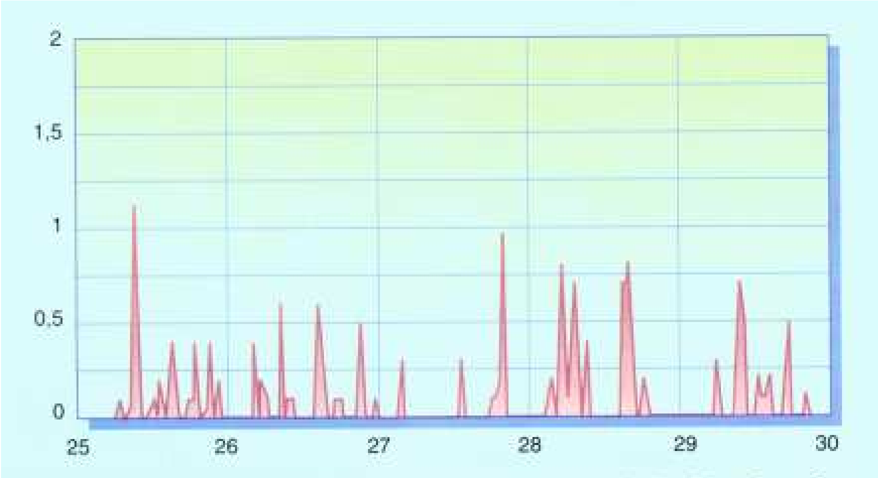


Figure 26 Observed undisturbed (Poisson) traffic, smoothed traffic and overflow

Hayward's approximation

Hayward's approximation provides an approximate way to calculate the blocking probability for non-Poissonian traffic (e.g. overflow traffic).

The starting point is the observation that for the occupancy N induced by Poissonian traffic in an infinite server system the following relations hold:

$$\boxed{\mathrm{E}[N] = a \quad \mathrm{V}[N] = a} \qquad N \sim \mathrm{Poisson}(a)$$

For non-Poissonian traffic, in contrast, we generally have $\mathrm{V}[N] \neq \mathrm{E}[N]$.

The Hayward approximation tries to describe non-Poissonian traffic by "equivalent Poisson traffic" and then apply Erlang's formula for the blocking probability.

The idea is to consider the behaviour of the occupied capacity R instead of the occupancy N.

Let

$$\begin{cases} c \;= & \text{the bandwidth (number of trunks) required of a single connection} \\ R \;=\; N \cdot c \;= & \text{the bandwidth occupied in the occupancy state } N \end{cases}$$

Hayward's approximation (continued)

For Poissonian traffic holds

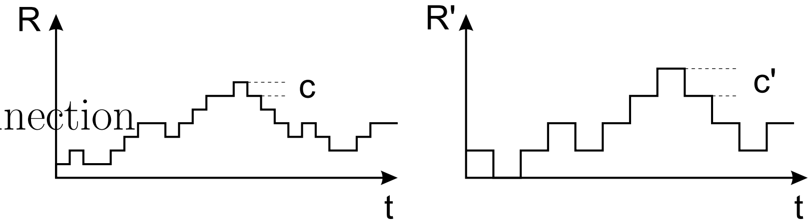
$$\begin{cases} \mathrm{E}[R] &= \mathrm{E}[c \cdot N] &= c \cdot a \\ \mathrm{V}[R] &= \mathrm{V}[c \cdot N] &= c^2 \cdot a \end{cases} \quad \Rightarrow \quad \frac{\mathrm{V}[R]}{\mathrm{E}[R]} = c$$

Consider a non-Poissonian source with known mean and variance of occupation

$$\begin{cases} \mathrm{E}[N] &= a \\ \mathrm{V}[N] &= v \end{cases} \quad \Rightarrow \quad \begin{cases} \mathrm{E}[R] &= c \cdot a \\ \mathrm{V}[R] &= c^2 \cdot v \end{cases}$$

This is now replaced by Poissonian traffic

$$\begin{cases} a' &= \text{intensity of the traffic} \\ c' &= \text{bandwidth requirement of a single connection} \end{cases}$$



so that the mean and variance of the occupied capacity are the same for the original non-Poisson traffic and the model Poisson traffic:

$$\mathrm{E}[R] = \mathrm{E}[R'], \quad \mathrm{V}[R] = \mathrm{V}[R']$$

The point is that for the equivalent Poisson traffic also the bandwidth required by a single connection is taken as free parameter.

A single connection of the equivalent traffic may thus require e.g. 1.6 trunks.

Hayward's approximation (continued)

The fitting of the two first moments leads to the conditions for a' and c'

$$\begin{cases} a' \cdot c' &= a \cdot c \\ a' \cdot c'^2 &= v \cdot c^2 \end{cases} \quad \Rightarrow \quad \boxed{\begin{cases} a' &= \dfrac{a^2}{v} \\ c' &= c \cdot \dfrac{v}{a} \end{cases}} \quad \begin{array}{l} \text{equivalent intensity} \\ \\ \text{equivalent bandwidth} \end{array}$$

The size of the system is modified correspondingly. If the original system has m trunks, i.e. a capacity of $m \cdot c$ bandwidth units, then it can accommodate $m \cdot c/c'$ equivalent connections. Thus the equivalent system has m' trunks:

$$m' = \frac{m \cdot c}{c'} = m \cdot \frac{a}{v}$$

Now the blocking probability is approximated by that of the equivalent Poissonian traffic:

$$\boxed{B \approx E(m', a') = E(m \cdot \frac{a}{v}, \frac{a^2}{v}) = E(\frac{m}{z}, \frac{a}{z})}$$

Hayward's approximation
where $z = v/a$

The load per the server is the same as before: $(a/z)/(m/z) = a/m$.
When $z > 1$, the system becomes smaller $\Rightarrow$ blocking increases.

Hayward's approximation (continued)

The non-Poissonian traffic may originate from several independent sources. If for each source the mean a_i and the variance v_i of the occupancy are known, then the corresponding parameters for the aggregate stream are

$$\begin{cases} a = \sum_i a_i \\ v = \sum_i v_i \end{cases}$$

Hayward's approximation then gives the approximate total blocking probability of the aggregate stream in a system with m trunks,

$$B \approx E(m \cdot \frac{a}{v}, \frac{a^2}{v}) = E(\frac{m}{z}, \frac{a}{z}) \qquad \text{where } z = v/a$$

How the blocking is distributed among different sources remains undefined.

The ERT method (Equivalent Random Theory)

The ERT method is also known as the Wilkinson method.

This provides another approximate method to calculate the blocking probability for non-Poissonian traffic.

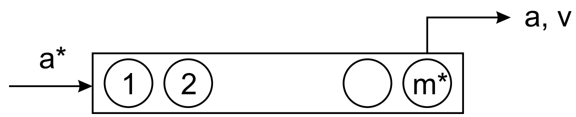
The offered traffic is characterized by

$$\begin{cases} a &= \text{traffic intensity} = \text{mean occupancy in an infinite system} \\ v &= \text{the variance of the occupancy in an infinite system} \end{cases}$$

In the case of several independent components we have

$$a = \sum_i a_i, \qquad v = \sum_i v_i$$

The idea of the ERT method is to think that the traffic (a, v) is obtained as overflow traffic from a fictitious channel



$$\begin{cases} a^* &= \text{offered traffic} \\ m^* &= \text{number of servers (trunks)} \\ & \quad\ \text{in the fictitious system} \end{cases}$$

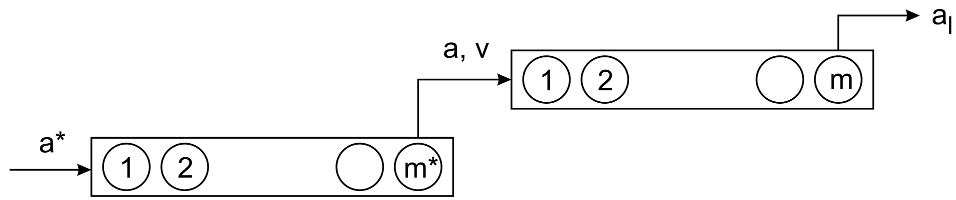
a^* and m^* are determined such that the overflow traffic in the fictitious channel has the intensity a and variance v.

The ERT method (continued)

Moment matching conditions are (the variance is given by the Riordan formula)

$$\begin{cases} a = a^* \cdot E(m^*, a^*) \\ v = a\left(1 - a + \dfrac{a^*}{m^* + 1 - a^* + a}\right) \end{cases} \quad \Rightarrow \quad a^*,\ m^*$$

If the channel to which the non-Poissonian traffic is offered has m trunks, the intensity a_ℓ of the ultimately overflown traffic can be calculated



$$a_\ell = a^* E(m^* + m, a^*)$$

An estimate for the traffic blocking is correspondingly

$$B = \frac{a_\ell}{a}$$

The ERT method (continued)

The solution to this pair of equations has to be found numerically.

An additional difficulty is that, in general, there is no solution for an integer m^*.

One has either to make further approximations or to extend the definition of Erlang's blocking formula for real valued trunk numbers (which, of course, are unrealistic).

Such an extension can indeed be made:

$$\boxed{E(m,a) = \frac{a^m e^{-a}}{\Gamma(m+1,a)}}$$

where the denominator is the incomplete gamma function

$$\Gamma(m+1,a) = \int_a^\infty t^m e^{-t} dt$$

By partial integration one can show for integer values of m that

$$\Gamma(m+1,a) = \int_a^\infty t^m e^{-t} dt = m! e^{-a} \left(1 + \frac{a}{1!} + \cdots + \frac{a^m}{m!}\right)$$

whereby the formula reduces to the familiar form

$$E(m,a) = \frac{\dfrac{a^m}{m!}}{1 + \dfrac{a}{1!} + \cdots + \dfrac{a^m}{m!}}$$

The ERT method (continued)

For the solution one can also use approximation formulae, e.g. that suggested by Rapp (1964),

$$\begin{cases} a^* &= v + 3z(z-1) \qquad\qquad \text{where } z = v/a \\ m^* &= \dfrac{a^*(a+z)}{a+z-1} - a - 1 \qquad \text{(this is an exact relation; approximation is in } a^*\text{)} \end{cases}$$

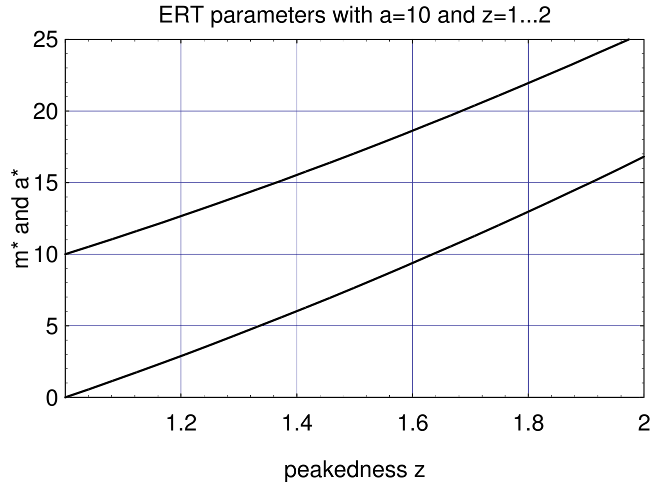
The approximation is not accurate if a is small and z is large.

The ERT method (continued)

The parameters of the ERT method

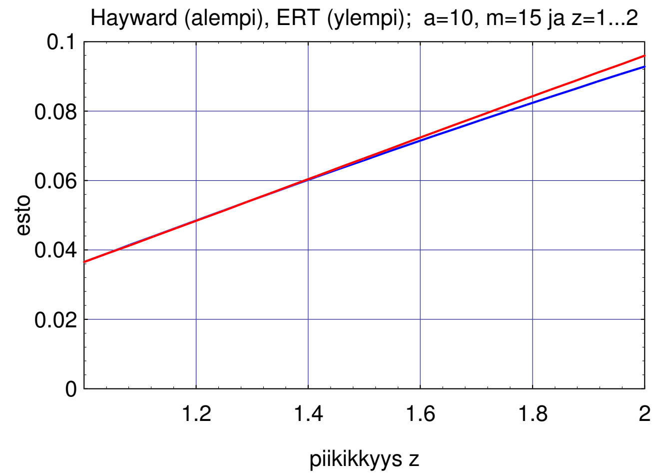
The parameters of the model system, m^* (lower curve) and a^* (upper curve) solved from the pair of equations are shown in the figure on the right, when the offered traffic has $a = 10$ and the peakedness $z = v/a$ varies from $z = 1 \ldots 2$.

In the solution we have used the exact extension of Erlang's function for real number of trunks.

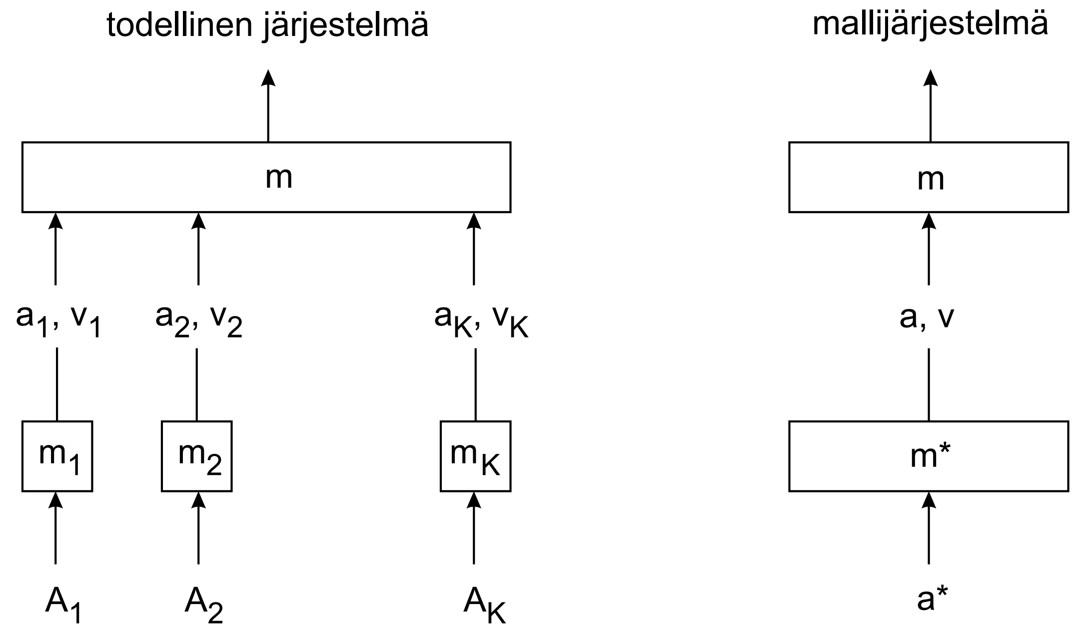


Comparison of the Hayward and ERT methods

The figure shows the blocking probability for traffic with $a = 10$ and $z = 1 \ldots 2$ offered to a $m = 15$ trunk system. The calculations have been done both with the Hayward method and the ERT method. In this case the results are very close to each other. One cannot say which one is "more correct". The blocking of the traffic, in reality, depends not only on the parameters a and $v = z \cdot a$ (and m). Based solely on these parameters the "correct result" in not known.



The ERT method (summary)



- The aggregate overflow traffic is handled as if it originated from a single channel.
- Find a^* and m^* such that the overflow traffic of the model system has the right a and v, $a = \sum_i a_i, \; v = \sum_i v_i$.
- The blocking in the overflow channel is $\dfrac{a^* E(m+m^*, a^*)}{a}$; total blocking $\dfrac{a^* E(m+m^*, a^*)}{\sum_i A_i}$