DIMENSIONING NETWORKS

1. Circuit switched networks

- Dimension the capacities (trunks/link speed) such that the total cost of the networks is minimized with the constraint that the end-to-end blocking probabilities may not exceed a given limit.
- Moe's principle

2. Packet switched networks

- "Square root method"
- Dimension the link capacities such that the mean delay of a packet is minimized given an upper bound to the total cost of the system.

On dimensioning circuit switched networks

The setting for the dimensioning problem is generally as follows. Given

- nodes and topology of the network (defining which nodes are connected by direct links)
- traffic matrix, which defines the offered traffic intensity $a_{i,j}$ from any source node i to any destination node j; the offered traffic is assumed to be Poissonian
- required grade of service, expressed as the maximum permissible blocking probability B; in principle, the allowed blocking probability $B_{i,j}$ may be specified separately for each origin-destination pair (i, j)
- link costs; depend on the capacity and the length of the link.

the task is to dimension the capacities C_j such that

- for all connections, the end-to-end blocking probabilities are below the given bound, and
- the cost of the network is minimized

Dimensioning problem

This is a typical problem of nonlinear optimization.

- The constraints depend nonlinearly on the capacities.
- The costs may depend nonlinearly on the capacities.

It should be noted that calculating the end-to-end blocking probabilities constitutes a problem itself. Usually, an exact calculation is not possible. The problem may be approached at three levels of accuracy:

1. The crudest approximation assumes that
   - the traffic offered to each link is Poissonian
   - the intensity of the offered traffic is the sum of the offered intensities of all the streams using the link
   - link blocking is given by the Erlang formula
   - the end-to-end blocking probability is the sum of the link blockings along the route
2. This can be improved by the so called reduced load approximation (to be presented later)
3. In some cases (e.g. hierarchical access network) an exact analysis is possible.

Dimensioning problem (continued)

We use the simplest approximation for estimating the end-to-end blocking probabilities.

In the case of a traditional circuit switched network, the problem is of the type of integer optimization, since the capacity (number of trunks) is a discrete variable.

By itself, solving a nonlinear optimization problem is a standard task and many program packages for that are available. In particular, there are specific programs for solving the network dimensioning problem.

In the following we consider the use of so called Moe's principle for finding an approximate solution for the optimization problem. [1]

[1] Our presentation follows the lecture notes by K. Kilkki

Moe's principle for a single link

Moe's principle defines an incremental approach by looking at the highest gain/cost ratio.

Consider first a single link with capacity n (trunks) and the offered traffic intensity a. If one more trunk is added, the intensity of the carried traffic is increased by the amount

$\Delta a = a(E(n,a) - E(n+1,a))$, where $E(n,a)$ is Erlang's blocking formula (B formula).

With the aid of this one can assess whether the gain obtained by the additional trunk is big enough to compensate for the costs.

- The number of trunks can be increased as far as the benefit exceeds the costs.
- With increasing number of trunks the marginal gain for each added trunk diminishes.

Moe's principle in a network

In the case of a network, one starts with a network which is underdimensioned, i.e. all the end-to-end blocking probabilities are greater than the allowed limit.

Then the capacity is increased trunk by trunk, by adding a trunk on the link where for a given cost the decrease in the end-to-end blocking probability is the greatest.

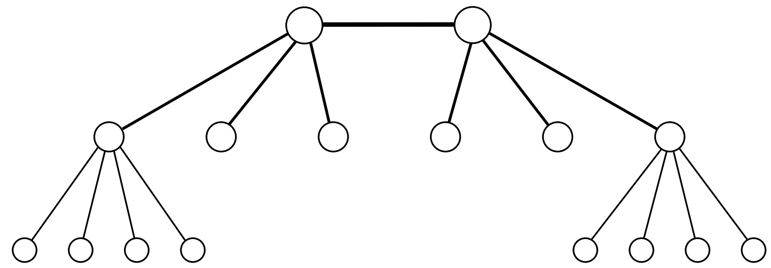
One continues this way until all the end-to-end blocking probability constraints are satisfied, i.e. are below the predefined limit B.

The problem is complicated by the fact that there are several connections going through a link and that the impact of incrementing the number of trunks on these connections is different. One may then consider the connection which the poorest service (greatest blocking) of those connections not meeting the constraint.

The incremental approach by Moe's principle does not necessarily lead to the global optimum but usually gives a pretty good solution.

Moe's principle in a symmetric hierarchical network

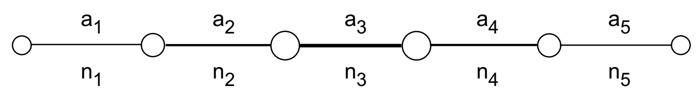
As an example, consider the hierarchical network of the figure. There are five levels of links in the network. Denote the number of links on hierarchy level i by m_i.



Assume that all links on a given hierarchy level are identical: the links have the same length and the offered traffic to all links is the same; on level i the offered traffic is a_i.

Then it is obvious that in an optimal dimensioning the capacities of all the links on a given level are equal, n_i.

The network can now be described by a simpler diagram.



Moe's principle in a symmetric hierarchical network (continued)

By Moe's principle, trunks are added to the level, where the obtained decrease in the blocking is greatest in relation to the cost. To this end we calculate the ratio

$$h_i = \frac{E(n_i, a_i) - E(n_i + 1, a_i)}{m_i c_i} , \qquad \text{where}$$

$$\begin{cases} n_i & = \text{ number of trunks in a single link on level } i \\ m_i & = \text{ number of links on level } i \\ c_i & = \text{ the cost of a single link on level } i \\ a_i & = \text{ intensity of the offered traffic to a link on level } i \end{cases}$$

Find level i with the greatest h_i and add one trunk to all the links on that level (i.e. m_i trunks in total).

Continue until the end-to-end blocking probability $\sum_i E(n_i, a_i)$ is smaller than B.

Dimensioning a symmetric hierarchical network: example

Optimize the network with the parameters given in the table.

The goal is get a end-to-end blocking probability less than 3%.

We start from an initial configuration given below left.

Greatest h_i is at level 3. It is advantageous to add trunks on that level until h_i for that level becomes smaller than the next highest h_i on level 2.

Continue the same way until the total blocking probability is less than 3 %. The final configuration is given in the table on the right.

i	m_i	c_i	a_i
1	50	100	3
2	5	200	25
3	1	300	100
4	10	200	12
5	100	100	2

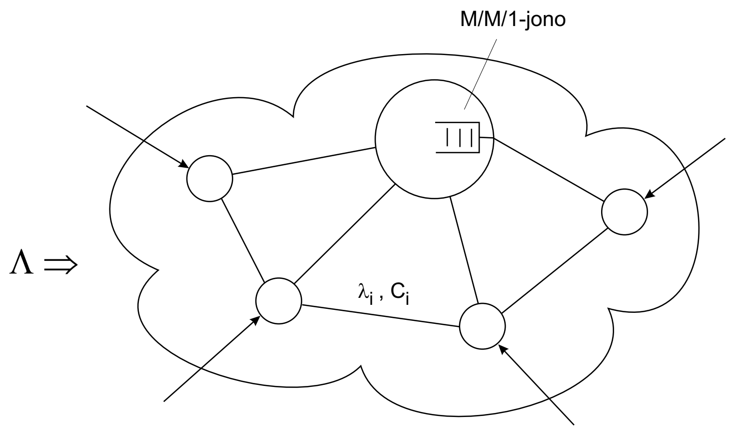
i	n_i	$B(n_i)$%	$B(n_i+1)$%	cost Meuro	extra cost Meuro	h_i %/Meuro
1	7	2.19	0.81	35	5	0.3
2	33	2.28	1.65	33	1	0.6
3	110	2.75	2.41	33	0.3	1.1
4	18	2.65	1.65	36	2	0.5
5	6	1.21	0.34	60	10	0.1
	total	11.08		197		

i	n_i	B_i %	cost Meuro
1	8	0.81	40
2	39	0.23	39
3	126	0.16	38
4	21	0.56	42
5	6	1.21	60
	total	2.96	219

In the final solution the high usage links have the least blocking probability.
Had the total blocking been distributed evenly between the levels, the total cost of the network would have been 230 Meuro.

Dimensioning a packet network ("square root method")

- There are M links, $i = 1, \ldots, M$, in the network.
- The size distribution of the packets is assumed to be exponential with mean size $1/\mu$ (bits)
- Model the output buffers of the nodes (routers) as $M/M/1$ queues.



$$\begin{cases} \lambda_i & = \text{packet stream (pckt/s) on link } i \\ C_i & = \text{capacity of link } i \\ d_i & = \text{specific price of link } i \text{ (euro/bps)} \\ 1/\mu & = \text{mean size of packet (bit)} \\ \Lambda & = \text{the total packet stream arriving to the whole network (packets/s)} \\ & \quad \text{note } \Lambda \neq \sum_i \lambda_i \end{cases}$$

The average sending time of a packet on link i is $1/(\mu C_i)$ $\Rightarrow$ the capacity of the link is μC_i (packets/s).

The mean sojourn time of a packet on link i (queueing + transmission)

$$\boxed{T_i = \frac{1}{\mu C_i - \lambda_i} = \frac{1}{\mu C_i} \cdot \frac{1}{1 - \lambda_i / \mu C_i}}$$

$\rho_i = \lambda_i / \mu C_i$ the load of link i

The objective and constraint of the optimization

One wishes to minimize the average time an arriving packet spends in the network

$$\boxed{T = \frac{1}{\Lambda}\sum_i \lambda_i T_i}$$

- This is a weighted sum of the sojourn times on different links.
- Can be interpreted by Little's result (the whole network is enclosed in the black box):

$$\begin{cases} \lambda_i T_i & = \text{ average number of packets (in queue + being transmitted) in buffer } i \\ \sum_i \lambda_i T_i & = \text{ average number of packets in the whole network} \\ \frac{1}{\Lambda}\sum_i \lambda_i T_i & = \text{ average sojourn time of packets in the network} \end{cases}$$

The constraint for the optimization is that the total cost may not exceed a given limit D,

$$\sum_i d_i C_i \le D$$

It is clear that the delay is minimized, if all the money is spent to increase the link capacities. Thus the unequality constraint can be replaced by an equality constraint

$$\boxed{\sum_i d_i C_i = D}$$

The task is to find capacities C_i such that under this constraint T is minimized.

Minimization under a constraint, the Lagrange multiplier method

The constraint of the minimization problem is taken into account by the method of the Lagrange multiplier

- To the objective function we add the function defining the constraint multiplied by a coefficient β which so far is undetermined. The function to be minimized is thus

$$G = \frac{1}{\Lambda}\sum_i \frac{\lambda_i}{\mu C_i - \lambda_i} + \beta\Big(\sum_i d_i C_i - D\Big)$$

- For any value β one can find the global (unconstrained) minimum of the function G.
- The location and the value of the minimum depend on the parameter β.
- Now, determine β such that the minimum satisfies the constraint, i.e. $\sum_i d_i C_i - D = 0$.
- Then the global minimum of the expression G is located on the hypersurface determined by the constraint.
- The restriction of function G on the hypersurface $\sum_i d_i C_i - D = 0$ is (irrespective of the value of β) the same as the original objective function.
- The global minimum of G is certainly also the minimum of the function G constrained on the hypersurface, i.e. the minimum of the original objective function on the hypersurface.
- Thus, the found solution minimizes the objective function and satisfies the constraint condition and is the solution for our problem.

Minimization

$$G = \frac{1}{\Lambda}\sum_i \frac{\lambda_i}{\mu C_i - \lambda_i} + \beta\Big(\sum_i d_i C_i - D\Big)$$

$$\frac{\partial G}{\partial C_i} = \frac{1}{\Lambda}\frac{-\lambda_i \mu}{(\mu C_i - \lambda_i)^2} + \beta d_i = 0, \qquad i = 1, \ldots, M$$

$$\Rightarrow (\mu C_i - \lambda_i)^2 = \frac{\lambda_i \mu}{\Lambda \beta d_i}$$

$$\Rightarrow C_i = \frac{\lambda_i}{\mu} + \frac{1}{\sqrt{\Lambda\beta\mu}}\sqrt{\frac{\lambda_i}{d_i}} \qquad \text{(minus sing is not possible)}$$

Now solve β from the requirement

$$D = \sum_i d_i C_i = \sum_i d_i \frac{\lambda_i}{\mu} + \frac{1}{\sqrt{\Lambda\beta\mu}}\sum_i \sqrt{\lambda_i d_i} \quad \Rightarrow \quad \frac{1}{\sqrt{\Lambda\beta\mu}} = \frac{D - \sum_i d_i \frac{\lambda_i}{\mu}}{\sum_i \sqrt{\lambda_i d_i}}$$

- λ_i/μ is the mean bit stream on link i; at least this capacity is needed on link i, in order to carry the load.
- Correspondingly, $\sum_i d_i \frac{\lambda_i}{\mu}$ is the minimum cost of the system.

Minimization (continued)

Denote

$$D_e = D - \sum_i d_i \frac{\lambda_i}{\mu}$$

This is the excess money left over for optimization, when the minimum capacities have been allocated.

In terms of D_e the constraint gives

$$\frac{1}{\sqrt{\Lambda \beta \mu}} = \frac{D_e}{\sum_i \sqrt{\lambda_i d_i}}$$

Substitute this back to the expression for the optimal C_i:

$$\boxed{C_i = \frac{\lambda_i}{\mu} + D_e \frac{\sqrt{\lambda_i / d_i}}{\sum_{j=1}^{M} \sqrt{\lambda_j d_j}}}$$

The link capacity exceeds the minimum value λ_i/μ (mean rate of the carried traffic [bits/s]) by an amount which is proportional to $\sqrt{\lambda_i/d_i}$.

The mean sojourn time in the optimized network is

$$\boxed{T_{\min} = \frac{1}{\Lambda \mu D_e} \Big(\sum_{i=1}^{M} \sqrt{\lambda_i d_i} \Big)^2}$$

Special case

If all the links have the same specific cost, one can set

$$\begin{cases} d_i &= 1 \\ D &= C \quad \text{the total available capacity} \end{cases}$$

Further, denote

$$\rho = \frac{1}{C} \sum_{i=1}^{M} \frac{\lambda_i}{\mu} \qquad \text{average utilization of the links}$$

Then the formulas take a simpler form:

$$\boxed{\begin{cases} C_i &= \dfrac{\lambda_i}{\mu} + (1-\rho)\, C\, \dfrac{\sqrt{\lambda_i}}{\sum\limits_j \sqrt{\lambda_j}} \\[2ex] T_{\min} &= \dfrac{\left(\sum\limits_i \sqrt{\lambda_i}\right)^2}{(1-\rho)\Lambda C \mu} \end{cases}}$$

The excess capacity is shared proportional to the $\sqrt{\lambda_i}$.