Fairness

Maxmin fairness

- An important consideration in 'best effort' type services
  - no quantitative QoS guarantees are given
  - all must receive service on a fair ground
- The *maxmin* definition of fairness

> A fair service maximizes the service of
> the customer receiving the poorest service

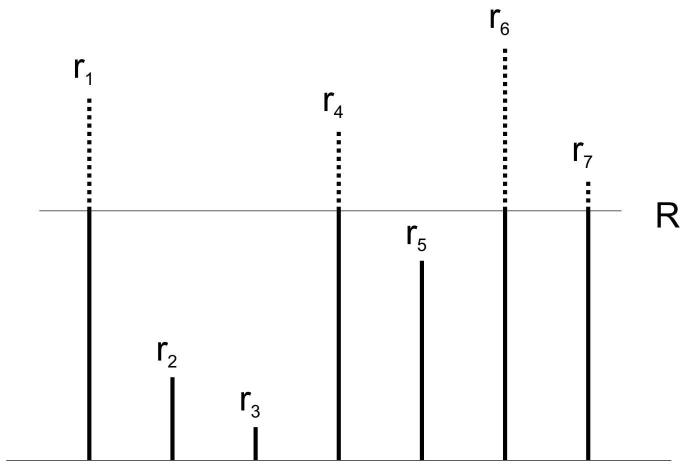
- In general, this does not define uniquely the resource sharing
  - if there are still some degrees of freedom left, one continues by maximizing the level of service of the customer receiving second poorest service etc.
- In the fair share each customer *either*
  - gets the service requested *or*
  - allocating more resource to the customer would worsen the service of some customer receiving the same level or poorer service
  - i.e. it is not possible to improve anybody's service only at the expense of customers receiving better service

Maxmin fairness (continued)

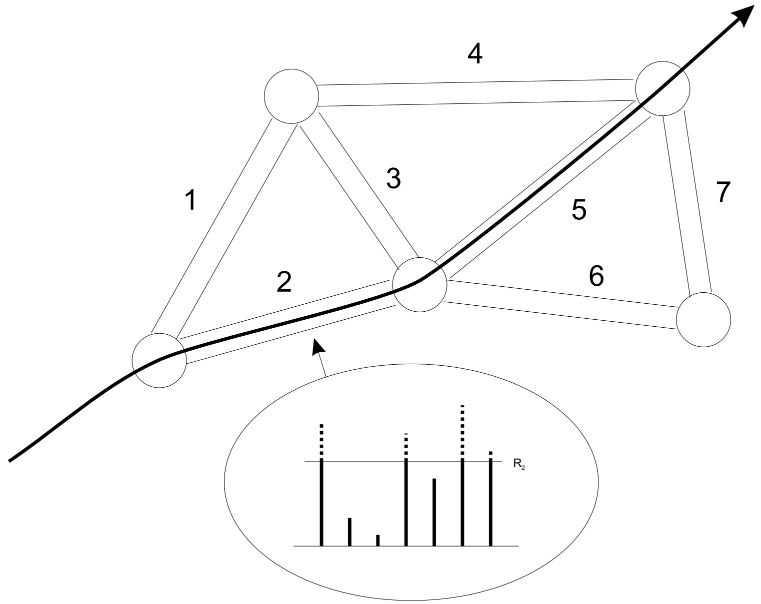
- Consider connections for which link ℓ is the bottleneck link
  - in case of fair share, the rates of these connections are equal
  - otherwise, the rate of the slowest connection could be increased by giving it more bandwidth from the faster connections
  - they have a common 'roof' R_ℓ

$$\sum_{i \in \mathcal{S}_\ell} \min(s_i, R_\ell) = C_\ell$$

$$\begin{aligned} \mathcal{S}_\ell &= \text{the set of connections} \\ &\quad\ \text{which use link } \ell \\ s_i &= \text{rate of source } i \end{aligned}$$



Maxmin fairness in a multinode network



$$\begin{cases} \sum_{i \in \mathcal{S}_\ell} \min(r_i^{(\ell)}, R_\ell) = C_\ell, \ \forall \ell \\ \\ r_i^{(\ell)} = \min_{\ell' \in \mathcal{L}_i - \{\ell\}} (R_{\ell'}, s_i) \end{cases}$$

$$\begin{cases} \mathcal{S}_\ell = & \text{the set of connections that use link } \ell \\ \mathcal{L}_i = & \text{the set of links used by connection } i \\ r_i^{(\ell)} = & \text{the rate at which source } i \text{ `would like' to send on the link } \ell \end{cases}$$

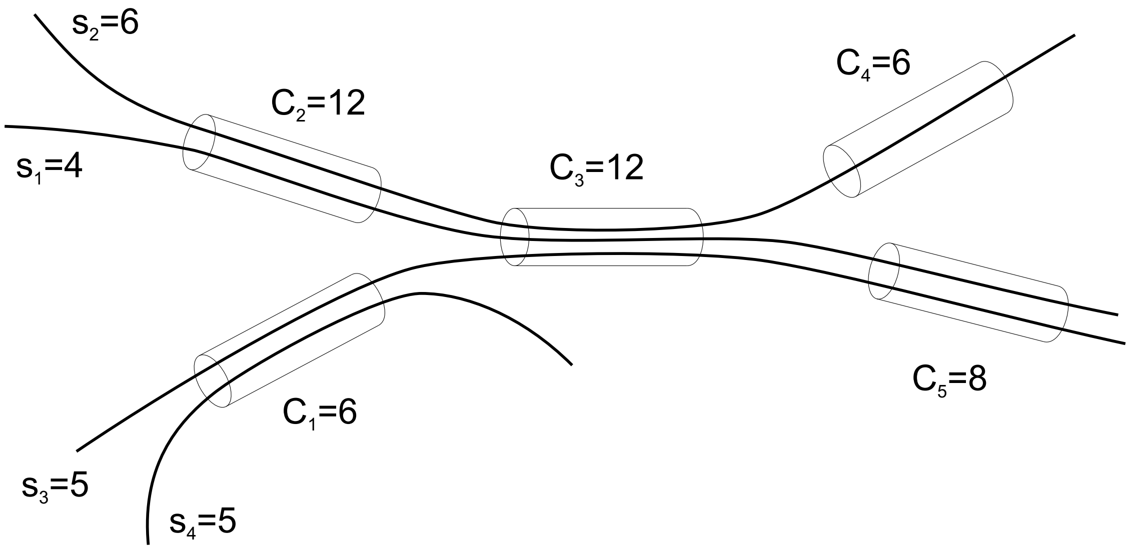
$$r_i = \min_{\ell' \in \mathcal{L}_i} (R_{\ell'}, s_i) = \min_{\ell' \in \mathcal{L}_i} r_i^{(\ell)} \quad \text{the rate allocated to source } i$$

Finding maxmin fair share (“filling algorithm”)

1. In the beginning set the rates of all connections to zero, $r_i = 0, \ \forall i$
2. Increase all rates (equally) until *either*
   - some of the sources has attained the requested rate *or*
   - the capacity of some link is fully used
3. ‘Freeze’ the rate of this connection / rates of these connections at the current level and continue increasing the rates of other connections as in point 2

- This algorithm requires centralized knowledge of the whole network
- There are also decentralized versions of this algorithm (e.g. for the ABR service in an ATM network), where the sources and switches exchange information
  - after a few iterations these algorithms converge to the fair share

Example of a maxmin fair share in a network



connection	1	2	3	4	Note
s_i	4	6	5	5	the requested rate
round					
0	0	0	0	0	
1	1	1	1	1	
2	2	2	2	2	
3	3	3	$\underline{3}$	$\underline{3}$	link 1 full
4	$\underline{4}$	4			requested rate of source 1
5		$\underline{5}$			link 3 full

- In this example the rate increase has been made in increments of one unit
- In reality, the increase must done in a continuous manner
  - it is easy to figure out, which limit is next encountered

Formal definition of maxmin fairness

The previous considerations can be set in a more precise mathematical form:

Definition 1: A rate vector of the connections $\mathbf{r} = \{r_i \mid i \in \mathcal{S}\}$ is feasible, if

$$\begin{aligned} 0 \le r_i \le s_i \quad \forall i \\ \sum_{i \in \mathcal{S}_\ell} r_i \le C_\ell \quad \forall \ell \end{aligned}$$

Definition 2: The rate vector $\mathbf{r}$ is maxmin fair, if it is feasible and if for each connection i and for each feasible rate vector $\hat{\mathbf{r}}$ for which $\hat{r}_i > r_i$, there is another connection j such that $r_j \le r_i$ and $\hat{r}_j < r_j$

Definition 3: For a given feasible rate vector $\mathbf{r}$ link ℓ is a bottleneck link for connection $i \in \mathcal{S}_\ell$, if $\sum_{k \in \mathcal{S}_\ell} r_k = C_\ell$ and $r_j \le r_i \ \forall j \in \mathcal{S}_\ell$

One can show that from these definitions it follows

Proposition 1: A feasible rate vector $\mathbf{r}$ is maxmin fair if and only if for each connection i some link is a bottleneck or $r_i = s_i$

Proposition 2: The maxmin fair rate vector $\mathbf{r}$ is unique

Utility based fairness definitions

- Maxmin fairness is the "classical" and best-known fairness concept
  - in the case of a single link an equal share of the bandwidth is obviously fair
  - in the network context a universal definition what is fair is far less obvious
  - maxmin fairness, while it can be arguably justified, is just one possible definition
- Other definitions have also been proposed
- So-called utility based fairness criteria encompass many possible definitions
  - maxmin fairness is a special case of utility based criteria
- The idea is to define a utility function $U(x_r)$ describing the utility a user (flow) on route r gets from the network if his capacity share is x_r
- The objective then is to maximize the total utility of all the users

$$U = \sum_{r \in \mathcal{R}} n_r U(x_r)$$

  where $\mathcal{R}$ is the set of all routes and n_r is the number of users (flows) on route r

Utility based fairness definitions (continued)

- Fair capacity sharing according to the utility criterion can now be defined as the solution of the optimization problem

$$\begin{cases} \text{max} & \sum_{r\in\mathcal{R}} n_r U(x_r) \\ \text{subject to} & Ax \le C \\ \text{over} & x \ge 0. \end{cases}$$

where

$$\begin{array}{lcll} x & = & \{x_r, r \in \mathcal{R}\} & \text{the vector of flow numbers on different routes} \\ C & = & \{C_j, j \in \mathcal{J}\} & \text{the vector of link capacities,} \\ A & = & \{a_{jr}, j \in \mathcal{J}, r \in \mathcal{R}\} & \text{the link-route incidence matrix;} \\ & & & a_{jr} \text{ is equal to } n_r \text{ if route } r \text{ uses link } j \text{ and 0 otherwise.} \end{array}$$

Utility based fairness definitions (continued)

- A reasonable and rather general choice for the utility function is

$$U(x) = \frac{x^{1-\alpha}}{1-\alpha}$$

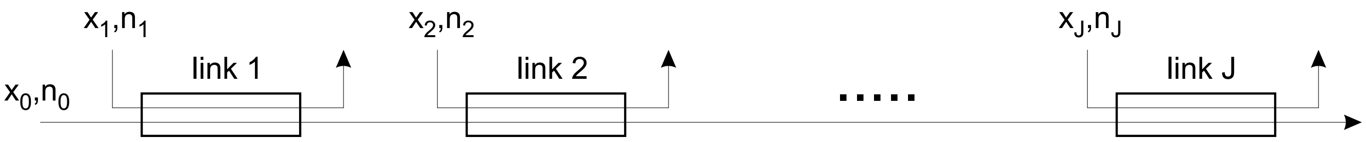
  where α is a free parameter

- Specific choices for α lead to the following important special cases

α	concept	$\max\limits_x \sum\limits_{\mathcal{R}} n_r U_r(x_r)$
0	maximize overall throughput	$\max\limits_x \sum\limits_{\mathcal{R}} n_r x_r$
1	proportional fairness	$\max\limits_x \sum\limits_{\mathcal{R}} n_r \log x_r$
2	minimize potential delay	$\min\limits_x \sum\limits_{\mathcal{R}} \frac{n_r}{x_r}$
∞	max min fairness	$\max\limits_x \min\limits_{r\in\mathcal{R}} x_r$

- Small values of α favour the common (network) utility at the expense of individuals; larger values of α emphasize the fairness towards the poorest guy.

Example: Utility based fairness in linear network



- The flow on the long route is indexed by 0; the flows on the short routes are indexed by the respective link number; all links have capacity 1

α	concept	x_0
0	maximize overall throughput	0
1	proportional fairness	$\dfrac{1}{n_0 + \Sigma_j n_j}$
2	minimize potential delay	$\dfrac{1}{n_0 + \sqrt{\Sigma_j n_j^2}}$
∞	maxmin fairness	$\dfrac{1}{n_0 + \max_{j \geq 1} n_j}$

- As α increases from 0 to ∞, the different allocations give relatively more bandwidth to long routes