EE 384Y Project

Final Report

Demand Based Rate Allocation

Arpita Ghosh and James Mammen
{arpitag, jmammen}@stanford.edu

June 6, 2003

Contents

Abstract

We consider the problem of fairly allocating rates in a network to a group of users based on the payments made by them, in a decentralized framework. We provide insight into the existing notion of proportional fairness by relating it to a method of rate allocation. We prove this for the case of two users and conjecture it to be true in general. We introduce the notion weighted max-min fairness for a network and propose a decentralized algorithm which we show to result in the weighted max-min fair allocation. We also provide a corresponding insight into what weighted max-min fairness means in terms of a method of rate allocation.

1 Motivation

In the present Internet, rate allocation is determined by the TCP algorithm. Depending on the load on a buffer at a router (at a link on the path from source to destination), which in turn is determined by the traffic of all active users, packets are dropped or marked. This feedback is used by the sender to modify its rate. Thus the rate achieved by a user depends on the aggregate traffic demand. This kind of scheme does not allow a rate allocation based on the amount a user is willing to pay to use the network. The average utilization of the network capacity may also be low when the number of users is large.

An interesting question is whether rate allocation and congestion control can be achieved by means of a supply-demand kind of mechanism, based on the amount each user is willing to pay. Given the amount each user is willing to pay and the route of each user, a centralized solution is certainly possible to compute. However a centralized solution is not desirable due to the following reasons. First, it would require a huge amount of communication between the the links in the network and the users. Also, if the topology of the network changes the solution will have to be recomputed. Most importantly, since the demands of the users are changing continuously such a solution is not very useful.

In a decentralized algorithm, each user would have to regulate its rate based on some feedback from the network which can be thought of as a price. Clearly the prices should reflect the load in the network, and so each link should generate a price based on the aggregate traffic passing through it. So such an algorithm would work by the users sending at some rate based on the current prices of the links and the links in turn modifying the prices to maximize the utilization and minimize congestion. We would like the distributed algorithm to be proportionally fair, lead to high resource utilization, prevent loss and congestion, and converge quickly to the equilibrium solution.

1.1 Fairness for a single link

Note that in the case of a single link of capacity C shared by N users, the fair allocation is quite clear. Each user i should be allocated a rate $x_i = (w_i / \sum w_j)C$. For a decentralized solution based on price, if the price were set to $p = \sum w_i / C$, then each user would get its fair rate by setting $x_i = w_i / p$. However in the case of a network with several links, what constitutes a "fair" allocation is not clear.

1.2 Previous work

Kelly, Maulloo and Tan [1] introduced the notion of proportional fairness for payment based rate allocation on a network. A feasible rate vector x is proportionally fair if for every other feasible rate vector y, the weighted aggregate of proportional changes is zero or negative:

$$\sum_i w_i \frac{y_i - x_i}{x_i} \leq 0.$$

They also proposed a decentralized algorithm, and showed that it converges to the proportionally fair allocation for the given network.

Mo and Walrand [2] generalized this to the notion of (α, p)-fairness: a feasible rate vector x is (α, p)-fair if for every other feasible rate vector y,

$$\sum_i w_i \frac{y_i - x_i}{x_i^\alpha} \leq 0.$$

This includes as special cases proportional fairness ($\alpha = 1$), and max-min fairness ($\alpha \to \infty$).

In the next section, we summarize our work. In section 3, we present two seemingly fair ways to allocate rates in a network and relate them to two notions of fairness. In Section 4 we introduce the concept of weighted max-min fairness, and present a decentralized algorithm which converges to the weighted max-min fair solution.

2 Summary Of New Results

The following, to the best of our knowledge, are the contributions of this project:

- We introduce a new notion of fairness for a network where users make payments which we call weighted max-min fairness. This is an extension of the classical notion of max-min fairness to the demand-based rate allocation scenario.
- The relation between a user's payment and the proportionally fair rate it is allocated is not clear, except in a mathematical sense. We present an intuitively fair method of rate allocation which results in a proportionally fair rate allocation. We can prove this for the case of 2 users and any network but conjecture it to be true for any number of users.
- As we did for proportional fairness, we present another intuitively fair method of rate allocation and prove that the rate allocation it results in is weighted max-min fair, this time for the case of any number of users and any network.
- We propose a decentralized algorithm which results in a weighted max-min fair rate allocation. We prove this using Lyapunov functions and the results presented in [2].

3 Network Of Resources With Multiple Users

Consider a network with L links and N users. Each user i has a route R_i which is a subset of these links. The routes are specified by a routing matrix $A \in \mathbf{R}^{n \times l}$, where $A_{ij} = 1$ if link j lies on user i's route, and $A_{ij} = 0$ otherwise. Suppose that each link j has capacity C_j, and each user i is willing to pay an amount w_i. The problem is to allocate the capacity on each link to the various flows through it, in a 'fair' manner, while ensuring that the utilization is also high. We consider two cases below depending on the way the payments are made by users.

3.1 Fixed link payments

Suppose the network will allocate rates on each link based on the *total* price w_i paid to it. Alternately, w_i could be the amount each user will pay per link it uses. In this case, a fair rate allocation $\{x_i : i = 1, \ldots, N\}$ must satisfy

$$x_i \geq \min_{l \in r_i} \left\{ \frac{C_j w_i}{\sum_{\{j:l \in r_j\}} w_j} \right\}, \quad i = 1, \ldots, N. \tag{1}$$

The fraction $\frac{C_j w_i}{\sum_{\{j:l \in r_j\}} w_j}$ is the minimum rate that should have been allocated to user i for (proportional) fairness if there were only link j in the user's route r_i. Since there are multiple links in r_i, the rate at which user i can actually send will be determined by the minimum of the above rates on each link on its route. In a decentralized setup, this would correspond to a user adjusting its rate based on the highest price on all links in its path.

However, every allocation satisfying (1) need not lead to a maximal utilization. We need to allocate rates in a fair manner and at the same time utilize the network resources to the maximum. Consider the following global iterative procedure for rate allocation. Let U be the set of users and J be the set of links.

Method 1

1. Every link allocates a rate to each user passing through it
$$x_i^j = \frac{w_i}{\sum_{s:j \in R_s} w_s} C_j.$$

2. Rate chosen by user i at the end of this round is the minimum of the above rates allocated to it on its links,
$$x_i' = \min_{j \in R_i} \{x_i^j\}.$$

3. If link j is completely utilized, *i.e.*,
$$y_j = \sum_{\{i:j \in R_i, i \in U\}} x_i' = C_j,$$
remove it from the network; also remove every user passing through this link from the set of users. For every link which has had a user removed from it, update the capacity of the link to be the capacity minus the sum of the rates of the removed users. This gives us a new (residual) network, with a residual (new) set of users U, links J, and the residual (new) capacities on these links.

4. Stop if there are no remaining users (*i.e.*, U is empty), else repeat method 1 for this residual network.

This procedure ends in a finite number of steps. At every step, at least one link is removed from the network for the following reason. Suppose, as for a single link, every link sets a price $p_j = \sum_{i:j \in R_i, i \in U} w_i / C_j$. The rate allocated to user i on link j, using the above procedure, is then $x_i^j = w_i / p_j$, and $x_i' = w_i / \max_{j \in R_i} p_j$. The link with the highest price will determine the x_i's for all the users passing through it and so this link will be completely utilized and hence removed.

It can be shown that this method of allocating rates will lead to a weighted max-min fair allocation (introduced in section 4). We have not included the proof since it is not central to the project. As an example, consider the network in Fig. 1 consisting of 3 users and 2 links. Users 1 and 2 pay 1 unit each, and user 3 pays 2 units. The above procedure results in the rate allocation $[1/3, 1/3, /2/3]$.

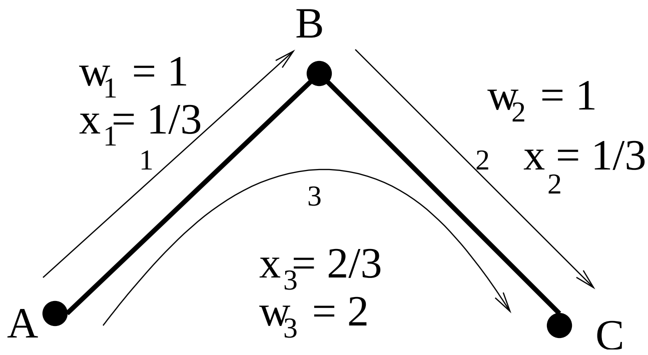

Figure 1: Example of rate allocation using Method 1

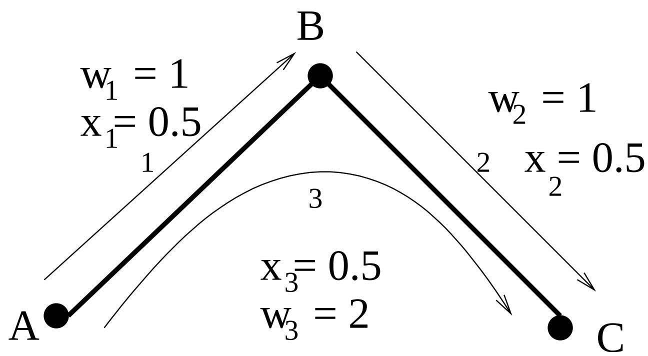

Figure 2: Example of rate allocation with fixed total payment to network

3.2 Fixed total payment to network

Next consider the case where each user i has a fixed amount w_i to spend across all the links in its route. User i splits the amount w_i into w_{ij}, on link $j = 1, \ldots, L$ so as to maximize its rate x_i. Rates are allocated on each link j in a proportionally fair manner using the w_{ij}s of the users i passing through it. The actual rate at which a user can send is the minimum of the rates allocated on all links on its route. Thus, in order to maximize its allocated rate, each user i should split its w_i so as to maximize the minimum of the rates allocated on all links on its route.

Consider the following iterative procedure for rate allocation.

Method 2

Let c_i^{min} be the minimum of the capacities of the links used by user i only. If user i has no such link then $c_i^{min} = \infty$. Let n_i be the number of links in user i's route. Initialize $w_{ij} = w_i/n_i$ for $j \in R_i$ and 0 otherwise.

1. Based on the w_{ij}s of all the users in the previous stage, each user solves for its w_{ij} so as to maximize its allocated rate x_i.

2. If $x_i > c_i^{min}$, solve again to get $x_i = c_i^{min}$.

3. If the allocated rate vector x is the same as at the previous iteration, stop. Or else, repeat.

As an example, consider the same network as before. This time the fair rate allocation turns out to be $[1/2, 1/2, 1/2]$ as shown in Fig. 2.

It can be shown that allocating rates using method 2 leads to a proportionally fair solution for the case of any network with two users. We do not include this proof here as it does not convey much insight into the general case. Based on several examples, we conjecture that this also holds for the general case of N users.

This provides insight into a proportionally fair rate allocation. Here a user's allocated rate x_i is proportional not only to the amount w_i it pays, but also depends on the number of links in its route. This contrasts with the previous case where the user's allocated rate x_i depends only on the amount he pays, and not the total resource of the network (in terms of capacity on all its links) that it consumes.

4 Payment-Based Max-Min Fairness

Recall the classical notion of max-min fairness: a feasible rate vector x is max-min fair if no rate x_i can be increased without simultaneously decreasing some x_j which is already less than or equal to x_i. This definition does not take into account the payments made by the users.

4.1 Weighted max-min fairness

In order to have a payment-dependent version of the above, we introduce the notion of weighted max-min fairness. We define a feasible rate vector x to be weighted max-min fair if no rate x_i can be increased without decreasing some x_j for which $x_j/w_j \leq x_i/w_i$.

This definition clearly includes the payments made by the users. Here are a few interpretations justifying this definition:

- A rate allocation x is weighted max-min fair if the rate for a user cannot be increased without decreasing the rate for some other user who is already paying as much or more per unit rate.
- A rate vector x is weighted max-min fair if the corresponding vector with x_i replaced by x_i/w_i is max-min fair in the classical sense, *i.e.*, weighted max-min fair for the vector of rates is max-min fair for the vector of rates per unit payment.
- If the payments w_i are all integers, weighted max-min fairness can be thought of as max-min fairness for the set of flows where user i's single flow is replaced by w_i different flows, each with rate x_i.

4.2 Decentralized algorithm

We would now like to come up with a decentralized algorithm that achieves the weighted max-min fair allocation of rates. In order to motivate the algorithm which follows, let us start with an example. Consider the two link, three user network in Fig3, where user 1 makes a payment of 1 unit, and user 2 and user 3 each make a payment of 2 units. Link AB allocates a rate of 1/3 to user 1 and 2/3 to user 2; link BC allocates a rate of 1/2 each to both users 2 and 3. User 3 (who is the only user using more than one link) can only send at rate 1/2. Here, the weighted max-min fair allocation of rates is $(1/2, 1/2, 1/2)$. Clearly, the rate of user 3 was constrained by link BC; in a decentralized algorithm, with prices set as in 1.1, the rate is determined by the highest priced link.

Consider now the following decentralized algorithm, with slotted time. User i adjusts its rate based on the highest price of all prices in its path:

$$x_i(n+1) = x_i(n) + \kappa[w_i - x_i(n) \max_{j \in R_i} p_j(n)]. \tag{2}$$

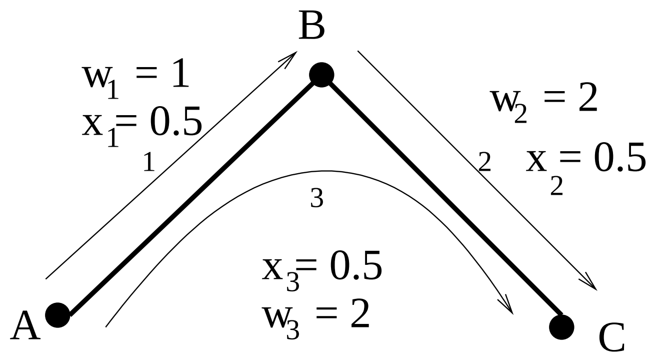

Figure 3: Example of rate allocation with fixed total payment to network

The price p_j is an increasing non-negative continuous function of the total traffic through link j:

$$p_j(n) = f_j\left(\sum_{i:j\in R_i} x_i(n)\right). \tag{3}$$

We make the additional, but completely unrestrictive assumption that f is not identically zero.

4.3 Proof of convergence

To analyze this decentralized algorithm, consider the differential equation corresponding to (2)-(3),

$$\frac{d}{dt}x_i(t) = \kappa\left(w_i - x_i(t)\max_{j\in R_i}\{p_j(t)\}\right) \quad i = 1,\ldots,N \tag{4}$$

$$p_j(t) = f_j\left(\sum_{i:j\in R_i} x_i(t)\right) \quad j = 1,\ldots,L \tag{5}$$

Using the fact that for a vector $a = (a_1,\ldots,a_n)$, $(\sum a_i^m)^{1/m} \to \max(a_i)$ as $m \to \infty$, we have the following series of approximations to (4):

$$\frac{d}{dt}x_i(t) = \kappa\left(w_i - x_i(t)(\sum_{j\in R_i} p_j^m(t))^{1/m}\right) \quad i = 1,\ldots,N \tag{6}$$

$$p_j(t) = f_j\left(\sum_{i:j\in R_i} x_i(t)\right) \quad j = 1,\ldots,L. \tag{7}$$

We will prove that (4) converges to the weighted max-min fair solution. There are two things to be proved: first, that the trajectory of (4) always converges to a unique solution, and second, that this solution is indeed the weighted max-min fair allocation of rates. To achieve this, we will first construct a Lyanpunov function which proves global stability of (6)-(7) for every m. Having done this, we will show that the fixed point of the Lyapunov function (to which all trajectories converge) can be made arbitrarily close to the optimal value of a certain optimization problem. As the parameter m tends to infinity, we will show that the solution to this optimization problem is indeed the weighted max-min fair solution.

Consider the following function

$$L(x) = -\frac{1}{m-1}\sum_i \frac{w_i^m}{x_i^{m-1}} - \sum_j \int_0^{\sum_{s:j\in R_s} x_s} p_j(y)^m dy \tag{8}$$

Theorem 1 *The strictly concave function $L(x)$ is a Lyapunov function for the system of differential equations (6)-(7). The unique value x maximizing $L(x)$ is a stable point of the system to which all trajectories converge.*

Proof. By our assumptions on the f_js (continuous, increasing, non-negative and not identically zero), and since $w_i > 0$ for every user i, we have that $L(x)$ is a strictly concave function on $x \geq 0$ with an interior maximum; which is therefore unique. To identify this maximum value, set the derivatives wrt the x_i to zero:

$$\frac{\partial}{\partial x_i} L(x) = \left(\frac{w_i}{x_i}\right)^m - \sum_{j\in R_i} p_j\left(\sum_{s:j\in R_s} x_s\right). \tag{9}$$

Further

$$\begin{aligned}\frac{d}{dt}L(x(t)) &= \sum_i \frac{\partial L}{\partial x_i}.\frac{d}{dt}x_i(t)\\ &= \kappa\left(\sum_i\left(\frac{w_i}{x_i}\right)^m - \sum_j p_j^m\left(\sum_{s:j\in R_s} x_s(t)\right)\right).\left(w_i - x_i(t)(\sum_{j\in R_i} p_j^m(t))^{1/m}\right)\\ &\geq 0.\end{aligned}$$

This shows that $L(x(t))$ is strictly increasing with t, unless $x(t) = x^*$, where x^* is the unique maximizing value of $L(x)$. So the function $L(x)$ is a Lyapunov function for the system of differential equations (6)-(7), and the theorem follows.

Choose the continuous function f_j, $j = 1, \ldots, L$, as

$$f_j(y) = (y - C_j + \epsilon)^+/\epsilon^2$$

where $(z)^+ = \max\{z, 0\}$. This function is clearly non-negative, increasing in y, and not identically zero. With this choice of f_j, it is clear that the maximizing value of (8) can be made arbitrarily close to the solution of the following convex optimization problem as $\epsilon \to 0$:

$$\begin{aligned}\max. \quad & \frac{-1}{m-1}\sum_i \frac{w_i^m}{x_i^{m-1}}\\ s.t. \quad & Ax \leq C\\ & x \geq 0\end{aligned} \tag{10}$$

From Lemmas 3.1 and 3.2 in [2], we know that the following optimization problem with $f_m(x) = -\frac{1}{x^m}$ results in a max-min fair solution as $m \to \infty$:

$$\begin{aligned}\max. \quad & \sum_i w_i f_m(x_i)\\ s.t. \quad & Ax \leq C\\ & x \geq 0\end{aligned} \tag{11}$$

Max-min fairness of the optimizing x is due to the following property of function f_m of giving more priority to smaller flows as m increases:

$$\frac{f'_m(x)}{f'_m(x+\epsilon)} \to \infty \text{ as } m \to \infty.$$

The cost function in (10) can be written as $\sum_i w_i f_m(\frac{x_i}{w_i})$ and it follows that the maximizing solution to (10) is max-min fair for the vector $(\frac{x_i}{w_i})$. From the definition of weighted max-min fairness, it is clear that the solution to (10) is weighted max-min fair.

Thus we have shown that all trajectories of the system of differential equations (6)-(7) converge to the unique weighted max-min fair rate allocation.

5 Conclusion

We have introduced the notion of weighted max-min fairness by extending the notion of max-min fairness to take into account the payments made by users. We proposed a decentralized algorithm which we analytically showed to result in the weighted max-min fair rate allocation. We also presented two intuitively fair methods of rate allocation and linked it with proportional and weighted max-min fairness.

References

[1] F. P. Kelly, A. K. Maulloo and D.K.H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability", *Journal of the Operational Research Society* 49, 1998.

[2] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control", *IEEE/ACM Trans. on Networking* 8-5, pp. 556-567, 2000.

[3] S. H. Low, F. Paganini, J. C. Doyle, "Internet congestion conrol", *IEEE Control Systems Magazine*, Feb 2002.

[4] R. J. Gibbens and F. P. Kelly, "Resource pricing and the evolution of congestion control", *Automatica* 35, 1999.