Understanding XCP: Equilibrium and Fairness

Steven H. Low*, Lachlan L. H. Andrew†, Bartek P. Wydrowski*
*CS and EE, California Institute of Technology
†ARC Special Research Centre for Ultra-Broadband Information Networks, University of Melbourne,
an affiliated programme of the National ICT Australia.

***Abstract*— We prove that the XCP equilibrium solves a constrained max-min fairness problem by identifying it with the unique solution of a hierarchy of optimization problems, namely those solved by max-min fair allocation, but solved by XCP under an additional constraint. We describe an algorithm to compute this equilibrium and derive a lower and upper bound on link utilization. While XCP reduces to max-min allocation at a single link, in a network the additional constraint can cause a flow to receive an arbitrarily small fraction of its max-min allocation. We present simulation results to confirm our analytical findings.**

***Index Terms*— Mathematical programming / optimization, Flow control.**

I. Introduction

TCP congestion control [1] has prevented severe congestion while the Internet underwent explosive growth during the last decade. However, the algorithm has shown serious difficulties as the network continues to scale in size and capacity [2], [3]. This has motivated several recent enhancements [4–9]. (See [6] for extensive references.) Of these, XCP [9] has received much attention for grid computing networks such as the OptIPuter, where its need for explicit communication between the traffic sources and the network is less of a deployment barrier than in the current Internet. Unlike most proposals, which set the flow rates according to the *sum* of congestion measures at the links of their paths, XCP sets them according to the *minimum* "available capacity" in their paths. This has the same flavor as MaxNet [10], [11]. XCP has been shown [9] to be stable when all round trip times (RTTs) are equal; however, no other analytic results are known. In this paper, we reverse engineer XCP to understand its equilibrium properties.

A deterministic fluid model of a general XCP network with multiple links and multiple flows is presented in Section II. Section III analyzes the equilibrium rates of XCP, and shows that all queues are empty in equilibrium. We prove the existence and uniqueness of XCP equilibrium rates by identifying them with the unique solution to a hierarchy of optimization problems. This is the same set of problems solved by the standard max-min fair allocation, but XCP solves them under an additional constraint. While XCP reduces to max-min allocation at a single link, its behavior in a network can be very different. We describe an algorithm to compute this equilibrium and derive bounds on link utilization.

In Section IV, we use these bounds to investigate the impact of the choice of protocol parameters on link utilization under the additional constraint. We show that flows can receive an arbitrarily small fraction of their max-min fair allocations. Specifically, with a max-min fair allocation, as long as a link is a bottleneck for some (not necessarily all) flows that pass through it, it will be fully utilized. Under XCP, this is no longer true: when the majority of flows using a link are bottlenecked at other links, the remaining flows at that link may not make efficient use of the residual bandwidth. With the parameters suggested in [9] however link utilization is at least 80% at any link. XCP has a "shuffling parameter" $\gamma > 0$ to prevent the network from settling into an unfair state [9]. We show that, given any network topology, we can choose γ sufficiently small so that the resulting allocation is close to max-min fairness. For any fixed $\gamma > 0$, however, there are topologies in which some flow rates can be far away from their max-min allocations.

These properties and the accuracy of our algorithm are verified by NS-2 simulations in Section V. We conclude in Section VI with limitations of this work. Some proofs are omitted due to space limitation but can be found in the full version on our web site [12].

Appears in Proc. of IEEE Infocom, Miami, FL, March 2005.

II. Model

Consider a network with L links shared by N flows. Sources are indexed by $i = 1, \ldots, N$, links by $l = 1, \ldots, L$ and packets by k. Let R be the $L \times N$ routing matrix: $R_{li} = 1$ if flow i uses link l and 0 otherwise. Let $L(i)$ be the set of links in the path of flow i:

$$L(i) \quad := \quad \{\ l \mid R_{li} = 1\ \}$$

and $I(l)$ be the set of flows that use link l:

$$I(l) \quad := \quad \{\ i \mid R_{li} = 1\ \}$$

Note that $l \in L(i) \Leftrightarrow i \in I(l)$.

We will present a continuous-time fluid model of XCP. For flows i, define the following variables:

- $w_i(t)$: window size at time t, in packets.
- τ_i: round-trip propagation (and fixed processing) delay.
- $T_i(t)$: round-trip time (RTT) at time t.
- $x_i(t) := w_i(t)/T_i(t)$: flow rate at time t.

For links l, define the following variables:

- c_l: capacity, in packets/sec.
- $b_l(t)$: backlog at time t, in packets.
- $y_l(t) := \sum_i R_{li} x_i(t)$: aggregate input rate at link l at time t. In equilibrium, we sometimes write $y_l(x)$ to emphasize the dependence on equilibrium rates x.

XCP divides time into control intervals of duration d, nominally the mean RTT of the flows at a link [9]. The RTT varies throughout the network and in time, and for simplicity we model d as a global constant.

To simplify notation, we assume all packets have the same size of 1 unit. We use "flow" and "source" interchangeably.

A. XCP description

We now summarize the XCP algorithm. See [9] for a detailed description. We do not model feedback delay because we consider only equilibrium properties in this paper.

For each packet, XCP generates a feedback signal prescribing a change in window size. Let $\tilde{H}_{lk}(t)$ be the feedback generated by link l for packet k at time t. The acknowledgment for packet k received by its source contains in its header the smallest feedback $\min_l H_{lk}(t)$ generated by links along its path. The source adds this quantity to its current window size.[1] We now describe how to compute the feedback.

Let

$$\phi_l(t) \quad = \quad \alpha d(c_l - y_l(t)) - \beta b_l(t)$$

where α, $\beta > 0$ are constants, c_l is the link capacity, $y_l(t)$ is the aggregate input rate, and $b_l(t)$ is the backlog at time t. Let $\phi_l^+(t) = \max(\phi_l(t), 0)$ and $\phi_l^-(t) = \max(-\phi_l(t), 0)$. The feedback on the kth packet at link l is

$$\tilde{H}_{lk}(t) \quad = \quad \tilde{p}_{lk}(t) - \tilde{n}_{lk}(t)$$

where $\tilde{p}_{lk}(t)$ and $\tilde{n}_{lk}(t)$ are the increase and decrease components respectively:

$$\tilde{p}_{lk}(t) \quad = \quad (h_l(t) + \phi_l^+(t))\frac{\tilde{T}_k(t)}{d}\frac{\tilde{T}_k(t)/\tilde{w}_k(t)}{\sum_{j=1}^{K_l(t)} \tilde{T}_j(t)/\tilde{w}_j(t)} \tag{1}$$

$$\tilde{n}_{lk}(t) \quad = \quad (h_l(t) + \phi_l^-(t))\frac{\tilde{T}_k(t)}{d}\frac{1}{K_l(t)} \tag{2}$$

where $\tilde{T}_k(t)$ and $\tilde{w}_k(t)$ are the round-trip time and window size, respectively, of the flow which transmitted packet k, and $K_l(t)$ is the total number of packets seen by link l over the time interval $(t - d, t]$. Here

$$h_l(t) \quad = \quad \max(0, \gamma d y_l(t) - |\phi_l(t)|)$$

is a "traffic shuffling" term with $\gamma \geq 0$ a constant. (Note that we are using the definition of γ from the appendix of [9], which differs by a factor of d from that used in the corresponding equation in [9].)

B. Dynamic model

We now translate the per-packet feedback $\tilde{H}_{lk}(t)$ into per-flow feedback. Let $H_{li}(t)$ be the feedback generated by link l for flow i at time t. In general, a quantity with a tilde ($\tilde{}$) pertains to a packet while the corresponding variable without a tilde pertains to a flow.

Substituting $\tilde{x}_k(t) = \tilde{w}_k(t)/\tilde{T}_k(t)$ in (1) gives

$$\tilde{p}_{lk}(t) \quad = \quad \frac{\tilde{T}_k(t)}{\tilde{x}_k(t)}\frac{h_l(t) + \phi_l^+(t)}{d\sum_{i=j}^{K_l(t)} 1/\tilde{x}_j(t)}. \tag{3}$$

$K_l(t)$ is the total number of packets arriving at link l in period $(t - d, t]$. For simplicity, we *assume* that

$$K_l(t) \quad = \quad y_l(t)d \ = \ d\sum_i R_{li} x_i(t)$$

[1] In practice, the window size has a lower bound of 1 packet, but for notational simplicity, we ignore this.

Of these packets, we *assume* that $R_{li}x_i(t)d$ packets are from flow i. Hence

$$\sum_{j=1}^{K_l(t)} \frac{1}{\tilde{x}_j(t)} \quad = \quad \sum_{i=1}^{N} R_{li}x_i(t)d \cdot \frac{1}{x_i(t)} \quad = \quad N_l d$$

Thus the per-packet feedback (3) becomes per-flow feedback

$$p_{li}(t) \quad = \quad \frac{T_i(t)}{d^2}\frac{h_l(t)+\phi_l^+(t)}{N_l x_i(t)}$$

Using $K_l(t) = y_l(t)d$ again, the per-packet feedback (2) becomes

$$n_{li}(t) \quad = \quad \frac{T_i(t)}{d^2}\frac{h_l(t)+\phi_l^-(t)}{y_l(t)}$$

The feedback *per packet* to flow i from link l is then

$$H_{li}(t) \quad = \quad \frac{T_i(t)}{d^2}\left(\frac{h_l(t)+\phi_l^+(t)}{N_l x_i(t)} - \frac{h_l(t)+\phi_l^-(t)}{y_l(t)}\right)$$

If flow i does not use link l, then set $H_{li}(t) = \infty$.

Let $H_i(t) = \min_{l\in L(i)} H_{li}(t)$ be the minimum feedback along i's path. Since source i receives $x_i(t)$ feedback packets per unit time (assuming every packet carries control information and is acknowledged), its window evolves according to:

$$\dot{w}_i(t) \quad = \quad x_i(t) \cdot H_i(t)$$

Substituting $x_i(t) = w_i(t)/T_i(t)$, we have

$$\dot{w}_i(t) = \frac{w_i(t)}{d^2}\min_{l\in L(i)}\left(\frac{h_l(t)+\phi_l^+(t)}{N_l x_i(t)} - \frac{h_l(t)+\phi_l^-(t)}{y_l(t)}\right)$$

Remark: The pseudo code in [9] contains additional "residual" terms. These use the feedback from upstream links to modulate the positive and negative components $\tilde{p}_{lk}(t)$ and $\tilde{n}_{lk}(t)$ to prevent excessive positive or negative feedback in each control period. However, it can be proved that the modulation of $\tilde{p}_{lk}(t)$ has no effect on the XCP equilibrium, and the modulation of $\tilde{n}_{lk}(t)$ also has no effect on the equilibrium if the average rate of flows bottlenecked at upstream links is significant (at least half that of flows bottlenecked at link l itself). Otherwise, the link utilization is slightly increased (by around 4% in Scenario 1 of Section V). Since these residual terms seem to impact primarily on dynamic rather than equilibrium properties, for simplicity, we ignore them in the analysis (but not the simulations) in this paper.

In summary, an XCP network is described by the following set of equations:

$$\dot{w}_i(t) \quad = \quad \frac{w_i(t)}{d^2}\min_{l\in L(i)} F_{li}(t) \tag{4a}$$

$$\dot{b}_l(t) \quad = \quad \begin{cases} y_l(t) - c_l & \text{if } b_l(t) > 0 \\ \max(y_l(t)-c_l, 0) & \text{if } b_l(t) = 0 \end{cases} \tag{4b}$$

where

$$F_{li}(t) \quad = \quad \frac{h_l(t)+\phi_l^+(t)}{N_l x_i(t)} - \frac{h_l(t)+\phi_l^-(t)}{y_l(t)} \tag{5a}$$

$$\phi_l(t) \quad = \quad \alpha d(c_l - y_l(t)) - \beta b_l(t) \tag{5b}$$

$$h_l(t) \quad = \quad \max(\gamma d y_l(t) - |\phi_l(t)|, 0) \tag{5c}$$

$$x_i(t) \quad = \quad \frac{w_i(t)}{T_i(t)} \tag{5d}$$

$$y_l(t) \quad = \quad \sum_i R_{li}x_i(t) \tag{5e}$$

$$T_i(t) \quad = \quad \tau_i + \sum_l R_{li}\frac{b_l(t)}{c_l} \tag{5f}$$

Here, $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$ are constants, and $\phi_l^+(t) = \max(\phi_l(t), 0)$, $\phi_l^-(t) = \max(-\phi_l(t), 0)$. Standard XCP uses $\alpha = 0.4$, $\beta = 0.226$ and $\gamma = 0.1$. We will study the behavior of the general model, which includes this as a special case. As we will see below, the qualitative properties, such as existence and uniqueness of equilibrium rates and their fairness properties, do not depend on specific values of these parameters (as long as $\gamma > 0$).

III. Equilibrium rates

This section characterizes the equilibrium of XCP and describes an algorithm to compute it; the next considers the implications of these results on utilization and fairness.

Equations (4)–(5) describe the evolution of the window vector $w(t) = (w_i(t), \text{ for all } i)$ and the backlog vector $b(t) = (b_l(t), \text{ for all } l)$. A pair of rate and backlog vectors (x, b), with window vector w given by $w_i = x_i(\tau_i + \sum_l R_{li}b_l/c_l)$, is said to be in *equilibrium* if both $\dot{w}(t) = 0$ and $\dot{b}(t) = 0$.

We start by defining a bottleneck link and other notation for XCP equilibrium. In general quantities without t dependence denote equilibrium quantities, e.g., w_i, T_i, x_i, F_{li}, etc.

Definition 1: A link l is said to be a bottleneck *for source i with respect to (w.r.t.) x if F_{li} is minimum among all the links that i uses, i.e., $F_{li} = \min_{m \in L(i)} F_{mi}$. In this case, source i is said to be* bottlenecked *at link l w.r.t. x.*

By definition, every source i has a bottleneck. Lemma 1 below implies that $F_{li} = 0$ in equilibrium at a bottleneck l.

We distinguish between links that are bottlenecks and those that are not. Let $L_1(i)$ be the set of links that are bottlenecks for source i w.r.t a given equilibrium rate x:

$$L_1(i) \;:=\; \{\, l \in L(i) \mid F_{li} = \min_{m \in L(i)} F_{mi} \,\}$$

and $L_0(i) := L(i) \setminus L_1(i)$. We also distinguish between sources that are bottleneck locally and those that are not. Let $I_1(l)$ be the set of sources bottlenecked at link l w.r.t. a given equilibrium rate x:

$$I_1(l) \;:=\; \{\, i \in I(l) \mid F_{li} = \min_{m \in L(i)} F_{mi} \,\}$$

and $I_0(l) := I(l) \setminus I_1(l)$. Let $N_l := |I(l)|$ be the number of sources at link l, $N_{l0} := |I_0(l)|$, and $N_{l1} := |I_1(l)|$. Let $\rho_l := N_{l0}/N_l$ be the fraction of flows through link l which are not bottlenecked at link l, and $\sigma_l := y_{l0}/c_l$ be the fraction of the link's capacity consumed by such flows. Note that while $L(i)$, $I(l)$, and N_l depend only on the routing matrix R, $L_1(i)$, $L_0(i)$, $I_1(l)$, $I_0(l)$, N_{l0}, N_{l1}, ρ_l and σ_l depend also on the equilibrium rate x through F_{li}.

From (4) and the definition of $I_0(l)$, we have

Lemma 1: The rate and backlog vector (x, b) is in equilibrium if and only if

1) *for all l, $y_l \le c_l$ with equality if $b_l > 0$ and*
2) *for all i, $\min_{l \in L(i)} F_{li} = 0$.*

Moreover,

3) *if $i \in I_0(l)$ and $j \in I_1(l)$ then $F_{li} > 0$ and $F_{lj} = 0$.*
4) *if $I_1(l) \neq \emptyset$ then $h_l = 0$ implies $\phi_l = 0$.*

Proof: Parts 1 to 3 are immediate. To see part 4, note that $h_l = 0$ implies

$$F_{li} \;=\; \frac{\phi_l^+}{N_l x_i} - \frac{\phi_l^-}{y_l}$$

By part 2, $F_{li} = 0$ for all $i \in I_1(l)$. Since at most one of ϕ_l^+ and ϕ_l^- can be nonzero, $\phi_l^+ = \phi_l^- = 0$, whence $\phi_l = 0$. ∎

A. The need for bandwidth shuffling

Without bandwidth shuffling, XCP would have $\gamma = 0$, giving $h_l(t) = 0$ for all l and t. In particular, $h_l = 0$ in equilibrium.

Theorem 2: Suppose $\gamma = 0$. Then (x, b) with $x_i = w_i/\tau_i$ is an equilibrium if and only if

1) *for all l, $y_l \le c_l$ and $b_l = 0$, and*
2) *for all i, there exists $l \in L(i)$ with $y_l = c_l$.*

Proof: The first condition in the theorem implies that for all l, $\phi_l \ge 0$. Combined with $h_l = 0$, this implies $F_{li} \ge 0$ for all i. The second condition then implies that for all i, $\min_{l \in L(i)} F_{li} = 0$. Hence, the conditions in the theorem are sufficient, by (4) and the first part of Lemma 1.

For necessity, there are two cases. If $I_1(l) \neq \emptyset$ then $\phi_l = 0$ by the second part of Lemma 1, and (5b) implies $y_l = c_l$ and $b_l = 0$, since $y_l \le c_l$, $b_l \ge 0$, and $\beta > 0$. Otherwise $l \in \bigcap_i L_0(i)$ and $F_{li} > 0$ by definition of $L_0(i)$. This implies $\phi_l^+ > 0$, and hence $y_l < c_l$, $b_l = 0$ in equilibrium. ∎

Remark: Without bandwidth shuffling, any (possibly unfair) boundary point of the set $\{x | Rx \le c\}$ would be an equilibrium. These are exactly the rates x which maximize aggregate throughput. This is why XCP uses $\gamma > 0$ [9].

The rest of the paper considers the more complicated case of $\gamma > 0$.

B. $\gamma > 0$ case: main results

This subsection provides a conceptually simple characterization and uses it to prove the existence and uniqueness of XCP equilibrium. In the next subsection, we provide an iterative algorithm to compute this equilibrium.

From (4)–(5) and Lemma 1, (x, b) is an XCP equilibrium if and only if

1) For all l, $y_l \le c_l$ with equality if $b_l > 0$.
2) For all sources i, $\min_{l \in L(i)} F_{li} = 0$.

Using (5a), condition 2 becomes: for all i, for all $l \in L(i)$,

$$x_i \;\le\; \frac{y_l}{N_l} \frac{h_l + \phi_l^+}{h_l + \phi_l^-} \;=:\; r^l \qquad (6)$$

with equality for some $l \in L(i)$. Hence for links l with $I_1(l) \neq \emptyset$, all flows $i \in I_1(l)$ that are bottlenecked at link l must have the *common* rate r^l. This has important implications as we will see below.

Several of the results will use the following technical lemma, whose proof is omitted.

Lemma 2: For all l

1) $x_i < x_j = r^l$ *if* $i \in I_0(l)$ *and* $j \in I_1(l)$.
2) $\sigma_l < \rho_l$ *if* $I_0(l) \neq \emptyset$.
3) $r^l \geq y_l/N_l$ *with equality if and only if* $I_0(l) = \emptyset$.
4) $h_l > 0$ *if* $I_1(l) \neq \emptyset$.
5) $y_l/c_l \geq \sigma_l$ *with equality if and only if* $I_1(l) = \emptyset$.

Unlike in the $\gamma = 0$ case, we characterize the equilibrium backlogs and rates separately. The following result says that the equilibrium queue under XCP is zero. This originates from the definition of ϕ_l in (5b), which is nonnegative in equilibrium. The same property is used in REM [13] to drive the queue to zero, or more generally, to a target value.

Theorem 3: In equilibrium, $b_l = 0$ and $\phi_l \geq 0$ for all l.

Proof: Links can be of three types: (a) $I_1(l) \neq \emptyset$, $I_0(l) = \emptyset$, (b) $I_1(l) = \emptyset$, $I_0(l) \neq \emptyset$, and (c) $I_1(l) \neq \emptyset$, $I_0(l) \neq \emptyset$. Each of these will be considered in turn.

Type (a) links are bottlenecks for all flows passing through them, i.e., links l where (6) holds with equality for all $i \in I(l)$. Since all flows have common rate r^l, $y_l = N_l r^l$, whence equality in (6) implies $\phi_l^+ = \phi_l^-$. Thus $\phi_l = 0$, and (5b) implies $y_l = c_l$ and $b_l = 0$, i.e., they share the link capacity fully and equally, with no queueing delay.

Type (b) links are not bottlenecks for any of the flows they carry. Hence, for all $i \in I(l)$,

$$x_i \;<\; \frac{y_l}{N_l}\frac{h_l + \phi_l^+}{h_l + \phi_l^-}$$

Multiplying both sides by R_{li} and summing over i, we have

$$y_l \;<\; \frac{y_l}{N_l}\frac{h_l + \phi_l^+}{h_l + \phi_l^-} \cdot \sum_i R_{li}$$

Hence

$$\frac{h_l + \phi_l^+}{h_l + \phi_l^-} \;>\; 1$$

Since both numerators and denominators are positive, $\phi_l^+ > \phi_l^-$. This implies $\phi_l > 0$ whence $y_l < c_l$ and $b_l = 0$.

Type (c) links are bottleneck links for some but not all of the flows using them. From (6), we have

$$\frac{h_l + \phi_l^+}{h_l + \phi_l^-} \;=\; \frac{r^l}{y_l/N_l} \;>\; 1$$

where the inequality follows from Lemma 2. As for type (b) links, this implies $\phi_l > 0$, $y_l < c_l$ and $b_l = 0$. ■

We next characterize the equilibrium rates of XCP. Define g_l as

$$g_l(x) \;:=\; \frac{\gamma y_l^2}{N_l\left[(\gamma+\alpha)y_l - \alpha c_l\right]}$$

where $y_l = \sum_i R_{li}x_i$. Since $g_l(x)$ depends on x only through y_l, we will abuse notation and also write $g_l(y_l)$ or $g_l(y_l(x))$. Define the *feasible set* of source rates x to be

$$X_0 := \left\{x \in \Re_+^N \,|\, g_l(y_l) < 0 \text{ or } x_i \leq g_l(y_l),\ \forall l, i \in I(l)\right\} \tag{7}$$

where $\Re_+$ denotes the set of nonnegative real numbers. We will later show that the XCP equilibrium must be in X_0. Note that $x \in X_0$ implies

$$Rx \;\leq\; c$$

To see this, multiply both sides of the inequality in (7) by R_{li} and sum over i to get

$$y_l \;=\; \sum_i R_{li}x_i \;\leq\; \frac{\gamma y_l^2}{(\gamma+\alpha)y_l - \alpha c_l}$$

Rearranging the above inequality yields $y_l \leq c_l$. The converse may not be true, i.e., X_0 may be a strict subset of $\{x|Rx \leq c\}$.

Our main result is to prove the existence and uniqueness of XCP equilibrium in a general network, and that this equilibrium solves a *constrained* max-min fairness problem.

Definition 4: A rate vector $x^* \in X_0$ is constrained max-min fair *if for any other feasible $x \in X_0$, $x_i > x_i^*$ implies that there is a j with $x_j < x_j^*$ and $x_j^* \leq x_i^*$.* Intuitively, a constrained max-min fair vector x^* is such that it is not possible to increase a component x_i^* without reducing another smaller or equal component x_j^*. This differs from standard max-min fairness only in that the feasible set X_0 is a subset of $\{x|Rx \leq c\}$ [14]. This restriction has important ramifications, as we will see in the next section.

We will prove constructively that the unique XCP equilibrium is constrained max-min fair by identifying it with the solution of a hierarchy of optimization problems over the feasible set X_0: it maximizes the smallest source rates in X_0, and then maximizes the second smallest rates over all rates that solve the first

problem, and so on. These maximization problems are defined inductively, following the idea of [15].

Let $L_0 = \emptyset$ and $I_0 = \emptyset$. The sets (L_0, I_0, X_0) define the first problem $\mathbf{P}_1$, whose solution is described by the sets (L_1, I_1, X_1). These sets in turn define the second problem $\mathbf{P}_2$, and so on. To simplify notation, let

$$\overline{L}_n := \bigcup_{m \le n} L_m \qquad \overline{I}_n := \bigcup_{m \le n} I_m$$

Given sets (X_0, L_0, I_0), ..., $(X_{n-1}, L_{n-1}, I_{n-1})$, if $\overline{I}_{n-1}$ contains all flows, then we stop. Otherwise, we define problem $\mathbf{P}_n$ and its solution L_n, I_n, X_n, $n \ge 1$, as follows.

$$\mathbf{P}_n\text{:} \qquad \max_{x \in X_{n-1}} \min_{i \notin \overline{I}_{n-1}} \quad x_i \tag{8}$$

Let

$$r_n := \min_{l \notin \overline{L}_{n-1}} \max_{x \in X_{n-1}} g_l(x) \tag{9}$$

$$L_n := \{ \text{ minimizing } l \text{ in (9) } \} \tag{10}$$

$$I_n := \bigcup_{l \in L_n} I(l) \setminus \overline{I}_n \tag{11}$$

$$X_n := \left\{ x \in X_{n-1} \,\middle|\, x_i \begin{cases} = r_n, & \forall i \in I_n \\ > r_n, & \forall i \notin \overline{I}_n \end{cases} \right\} \tag{12}$$

A few important properties are immediate from these definitions. First, the rates r_n are monotonic:

$$\min_l \frac{c_l}{N_l} \;=\; r_1 < r_2 < \cdots < r_n \tag{13}$$

Second, L_n and I_n are nonempty; moreover they are disjoint from $\overline{L}_{n-1}$ and $\overline{I}_{n-1}$, respectively. Hence $\overline{I}_n$ will eventually contain all the flows and there are only a finite number of problems $\mathbf{P}_n$. Finally, X_n are *strictly* nested:

$$X_0 \supsetneq X_1 \supsetneq \cdots \supsetneq X_n$$

Indeed it will become clear that X_n is exactly the set of solutions to problem $\mathbf{P}_n$, i.e., X_1 is the set of feasible rates $x \in X_0$ whose smallest rates are maximized, X_2 is a subset of X_1 whose second smallest rates are also maximized, and so on. We prove below that if $\mathbf{P}_{n^*}$ is the last problem, then X_{n^*} is a singleton that solves all problems $\mathbf{P}_1, \ldots, \mathbf{P}_{n^*}$.

To contrast XCP equilibrium with the standard max-min fair allocation, we derive a "bottleneck" characterization that is analogous to that for max-min fairness; see the beginning of Section IV.

Lemma 3: Suppose x is the XCP equilibrium rate vector. Link l is a bottleneck for source $i \in I(l)$ w.r.t. x if and only if

1) $x_i = g_l(x)$, *and*
2) $x_i \ge x_j$ *for all* $j \in I(l)$.

Proof: Suppose link l is a bottleneck link for source i w.r.t. equilibrium x. Then Lemma 1(2) implies that $F_{li} = 0$, i.e., equality holds in (6). Since $\phi_l \ge 0$ by Theorem 3 and $h_l > 0$ by Lemma 2, (5c) becomes $h_l = \gamma d y_l - \phi_l$. Thus from (6)

$$x_i \;=\; r^l \;=\; \frac{y_l}{N_l} \frac{\gamma d y_l}{\gamma d y_l - \phi_l} \;=\; g_l(x) \tag{14}$$

proving the first condition. Condition (6) then implies the second condition.

Conversely, suppose the two conditions are satisfied. If $h_l = 0$, then $F_{li} = 0$ from (5a). Lemma 1(2) then implies F_{li} is the minimum among links in source i's path, i.e., link l is a bottleneck. On the other hand, if $h_l > 0$, then, as above, $\phi_l \ge 0$ and $h_l = \gamma d y_l - \phi_l$. Then $x_i = g_l(x)$ is equivalent to $F_{li} = 0$, proving that l is a bottleneck. ■

Motivated by this lemma, we call link l a *nonbottleneck w.r.t.* x if either $g_l(x) < 0$ or $x_i < g_l(x)$ for all $i \in I(l)$.

Our main result is

Theorem 5: The problems $\mathbf{P}_n$ are well-defined and have a unique solution. Moreover, the following are equivalent:

1) *x^* is an XCP equilibrium.*
2) *x^* is the unique rate vector that solves all the problems $\mathbf{P}_n$.*
3) *x^* is constrained max-min fair.*
4) *$x^* \in X_0$ and every flow has a bottleneck w.r.t. x^*, i.e., for all i, there is an $l \in L(i)$ such that $x_i^* = g_l(x^*)$ and $x_i^* \ge x_j^*$ for all $j \in I(l)$.*

Before presenting our proof, we derive a (centralized) algorithm to compute the XCP equilibrium.

C. Algorithm for computing equilibrium

The equilibrium rates of XCP can be found using an algorithm analogous to that of [14] for max-min fairness. However, because the constraint on the link throughput in (6) depends on the aggregate flow rate through h_l and ϕ_l, some extra bookkeeping is required.

Theorem 6: The utilization of a bottleneck link l satisfies

$$\frac{y_l}{c_l} = \frac{\alpha + (\gamma+\alpha)\sigma_l + \sqrt{(\alpha-(\gamma+\alpha)\sigma_l)^2 + 4\alpha\gamma\sigma_l(1-\rho_l)}}{2(\gamma\rho_l+\alpha)} \tag{15}$$

The rates of all sources $i \in I_1(l)$ bottlenecked at l satisfy

$$r^l = \frac{c_l}{N_l}\frac{[\Xi_l - \gamma\sigma_l\rho_l] + \sqrt{[\Xi_l+\gamma\sigma_l\rho_l]^2 - 4\alpha\gamma\sigma_l(\rho_l-\sigma_l)}}{2(1-\rho_l)(\gamma\rho_l+\alpha)} \tag{16}$$

where

$$\Xi_l = (\gamma\sigma_l+\alpha)(1-\sigma_l) - \gamma\sigma_l(\rho_l-\sigma_l) \tag{17}$$

Proof: Substituting $r^l = (y_l - y_{l0})/(N_l - N_{l0})$ into (14) and solving the resulting quadratic equation gives

$$y_l = \frac{\alpha c_l + (\gamma+\alpha)y_{l0} \pm \sqrt{(\alpha c_l + (\gamma+\alpha)y_{l0})^2 - K}}{2(\gamma N_{l0}/N_l + \alpha)} \tag{18}$$

where $K = 4\alpha c_l y_{l0}(\gamma N_{l0}/N_l + \alpha)$. It can be proved that only the larger solution of (18) satisfies part 5 of Lemma 2, and is a valid equilibrium. Rearranging the term in the square root gives (15).

To obtain (16), instead substitute $y_l = (N_l - N_{l0})r^l + y_{l0}$ into (14), giving

$$A_l\left(\frac{r^l N_l}{c_l}\right)^2 + B_l\frac{r^l N_l}{c_l} + C_l = 0, \tag{19}$$

where

$$\begin{aligned} A_l &= (1-\rho_l)(\gamma\rho_l+\alpha) \\ B_l &= 2\gamma\sigma_l\rho_l - \gamma\sigma_l - \alpha(1-\sigma_l) \\ C_l &= -\gamma\sigma_l^2. \end{aligned}$$

Since y_l is increasing in r^l, it is again only the larger root which represents the XCP equilibrium. Thus

$$\frac{r^l N_l}{c_l} = \frac{[\Xi_l - \gamma\sigma_l\rho_l] + \sqrt{[\Xi_l-\gamma\sigma_l\rho_l]^2 - 4A_lC_l}}{2(1-\rho_l)(\gamma\rho_l+\alpha)},$$

where Ξ_l is given in (17). Rearranging the expression in the square root gives (16). ■

Note that the right-hand side of (16) depends on the rate vector x throughput ρ_l and σ_l. Hence it is not an explicit formula for the throughput of a general flow. However it says that the common "bottleneck" rate at each link l depends on the rate vector x only through y_{l0} and N_{l0} that are bottlenecked elsewhere. These are source rates smaller than the "bottleneck" rate at link l, by Lemma 1. This motivates an algorithm similar to the max-min algorithm of [14] that calculates the throughput x_i of each flow in increasing order, without the need for recourse to simulation.

1 Set $\bar{I}_0 \leftarrow \emptyset$, $\bar{L}_0 \leftarrow \emptyset$, $\sigma_l(0) \leftarrow 0$, $\rho_l(0) \leftarrow 0$ for all l, $n \leftarrow 1$
2 **repeat**
  2.1 For each link, $l \notin L_{n-1}$ find $r^l(n)$ from (16) using $\sigma_l(n-1)$ and $\rho_l(n-1)$ from rates already allocated
  2.2 Set $r_n \leftarrow \min_j r^j(n)$
  2.3 Set $L_n \leftarrow \{l : r^l(n) = r_n\}$
  2.4 **foreach** $l \in L_n$
    2.4.1 Set $r^l \leftarrow r_n$
    2.4.2 For each flow $i \in I(l) \setminus \bar{I}_n$, set $x_i \leftarrow r^l$
    endfor
  2.5 Set $I_n \leftarrow \bigcup_{j\in L_n} I(j) \setminus \bar{I}_n$
  2.6 Set $\bar{I}_n \leftarrow \bar{I}_{n-1} \cup I_n$
  2.7 Set $\bar{L}_n \leftarrow \bar{L}_{n-1} \cup L_n$
  2.8 **foreach** $l \in \bigcup_{i\notin\bar{I}_n} L(i)$
    2.8.1 Set $\sigma_l(n) \leftarrow \sigma_l(n-1) + \sum_{i\in I_n} R_{li}x_i/c_l$
    2.8.2 Set $\rho_l(n) \leftarrow \rho_l(n-1) + \sum_{i\in I_n} R_{li}/N_l$
    endfor
  2.9 Set $n \leftarrow n+1$
until $\bar{I}_n = \{\text{all flows}\}$

This solves each of the optimization problems, $\mathbf{P}_n$, in turn. The key is that, by keeping track of the used capacity of each link, $\sigma_l(n)$ and $\rho_l(n)$, it can compute the maximization in (9) in closed form. For each l, the values $\sigma_l(n)$ and $\rho_l(n)$ vary during the algorithm. For the algorithm to be correct, they must have the right values when link l is the minimum in step 2.2. This occurs because the link rates are allocated in increasing order [12].

If $\gamma = 0$ then (16) reduces to $r^l = (c_l - y_{l0})/(N_l - N_{l0})$, and hence the algorithm reduces to the algorithm in [14] to compute the max-min fair allocation. This suggests that, given any topology specified by the routing matrix R and link capacity vector c, one can choose $\gamma > 0$ to be sufficiently small so that the equilibrium of (4) is close to max-min fair. On the other hand, with small γ, the convergence of individual rates to fairness can be very slow. We will return to this point in Section IV.

D. $\gamma > 0$ case: proofs and intuitions

In this subsection we establish Theorem 5. For the complete proof, see [12]. We start with a simple observation that greatly simplifies the solution of $\mathbf{P}_n$.

Lemma 4: Suppose X_n is nonempty. The maximization in (9) can be taken over $x \in X_n$ that have equal x_i for $i \notin \overline{I}_n$.

In view of Lemma 4, we can replace X_n in (12), for $n \geq 1$, by their subsets:

$$\hat{X}_n := \left\{ x \in X_{n-1} \,\middle|\, x_i = \begin{cases} r_n, & \forall i \in I_n \\ r_n + \epsilon, & \forall i \notin \overline{I}_n, \epsilon > 0 \end{cases} \right\} \quad (20)$$

and use them instead of X_{n-1} in computing r_n:

$$r_n := \min_{l \notin \overline{L}_{n-1}} \max_{x \in \hat{X}_{n-1}} g_l(x)$$

This greatly reduces the complexity of (9) from maximizing over n-vectors $x \in X_n$ to over a scalar $\epsilon > 0$.

Denote an $x \in \hat{X}_n$ by $x(\epsilon; n)$, with

$$x_i(\epsilon; n) = \begin{cases} r_m, & i \in I_m, m \leq n \\ r_n + \epsilon, & i \notin \overline{I}_n \end{cases} \quad (21)$$

and let $x(0; n) := \lim_{\epsilon \to 0} x(\epsilon; n)$. Note that $x(0; n)$, $n \geq 1$, is not in X_n according to definition (12), though it is in X_0. We will see in Lemma 6 below that $x(0, n)$ plays an important role in the proof of Theorem 5. The vector $x(\epsilon; n)$ induces link flows

$$\begin{aligned} y_l(\epsilon; n) &= \sum_i R_{li} x_i(\epsilon; n) \\ &= \sum_{m=1}^{n} r_m \sum_{i \in I_m} R_{li} + r_n \sum_{i \notin \overline{I}_n} R_{li} + \epsilon \sum_{i \notin \overline{I}_n} R_{li} \quad (22) \end{aligned}$$

This motivates the following main technical lemma.

Lemma 5: Given any scalars $z \geq 0$, $\eta > 0$, and $s \geq 0$, define

$$\begin{aligned} \hat{g}_l(\epsilon) &:= \hat{g}_l(\epsilon; z, \eta) := g_l(z + \eta\epsilon) \\ &= \frac{\gamma(z + \eta\epsilon)^2}{N_l((\gamma + \alpha)(z + \eta\epsilon) - \alpha c_l)} \quad (23) \end{aligned}$$

for some $N_l \geq 1$, $\alpha > 0$, $\gamma > 0$ and $c_l > 0$.

1) *If either $\hat{g}_l(0) < 0$ or $s \leq \hat{g}_l(0)$ then there exists a unique $\epsilon_l \geq 0$ such that $s + \eta\epsilon_l = \hat{g}_l(\epsilon_l)$, where $\epsilon_l = 0$ if and only if $s = \hat{g}_l(0)$.*
2) *Moreover, over $\{\epsilon \mid \hat{g}_l(\epsilon) > 0\}$, $s + \eta\epsilon \leq \hat{g}_l(\epsilon)$ if and only if $\epsilon \leq \epsilon_l$.*

Lemma 5 implies that if link l is a bottleneck for some source i with respect to an $x \in X_0$, then the rate of source i cannot be increased without violating the feasibility constraint in (7). For instance, let $n \geq 1$ be such that $l \in L_n$. Setting $z = y_l(0; n-1)$, $s = r_{n-1}$ and $\eta = \sum_{i \notin \overline{I}_{n-1}} R_{li}$ gives $\epsilon_l = r_n - r_{n-1}$ and l is a bottleneck for all $i \in I_n$ w.r.t. $x(\epsilon_l; n-1)$. Lemma 5(b) then implies that rates greater than r_n are infeasible at link l.

The next lemma implies that all links $l \in L_n$ are bottlenecks w.r.t. all $x \in \hat{X}_n$, and all links $l \notin \overline{L}_n$ are nonbottlenecks w.r.t. $x(0; n)$. In particular, this implies that X_n are nonempty.

Lemma 6: For each $n \geq 1$,

1) *if $l \in L_n$, then $x_i = g_l(x)$ for all $i \in I_n$ w.r.t. all $x \in \hat{X}_n$.*
2) *if $l \notin \overline{L}_n$, then either $g_l(x(0; n)) < 0$ or $x_i(0; n) < g_l(x(0; n))$ for all $i \in I(l)$.*

A sketch of the proof of Theorem 5 is as follows. For details, see [12]. First note that the optimization problems are well defined and that characterizations 2 and 3 are equivalent.

Lemma 5 can be shown to imply the equivalence of characterizations 3 and 4 by the following contradiction argument based on that for standard max-min fairness [14]. Assume there is an unfair x^*, for which every flow has a bottleneck. Some flow, i, bottlenecked at l, can have its rate increased, giving rates x. There exist an $\hat{x}$ and ϵ such that (a) $\hat{x}_j = x^*_j + \epsilon$ if $j \in I(l)$ and $\hat{x}_j = x^*_j$ otherwise, (b) $y_l(x) = y_l(\hat{x})$, and (c) $\hat{x}_j \leq g_l(y_l(\hat{x}))$. Since l is a bottleneck, $x^*_i = g_l(y_l(x^*)) > 0$. Applying Lemma 5(1) with $z = y_l(x^*)$, $s = x^*_i$ and $\eta = \sum_j R_{lj}$ gives $\epsilon_l = 0$. Since $\hat{x}_j = x^*_j + \epsilon$ for all $j \in I(l)$, with $\epsilon > \epsilon_l$, Lemma 5(2) implies that $\hat{x}_i > g_l(y_l(\hat{x}))$, contradicting (c). Conversely, if source i has no bottleneck, then Lemma 5 gives an $\epsilon = \min_l(\epsilon_l) > 0$ by which rate i can be increased.

It remains to show that 1 and 4 are equivalent. The discussion at the beginning of Section III-B shows that x^* is an XCP equilibrium if and only if, for all i, (6) holds for all $l \in L(i)$, with equality for some $l \in L(i)$. This, with (14), establishes $x^* \in X_0$. As observed after Definition 1, every flow has a bottleneck by definition.

To show 4 implies 1, it suffices to show that the characterization in Lemma 3 implies statements 1. and 2. at the start of Section III-B. The discussion after (7), and setting $b = 0$, establishes Statement 1. This shows $\phi_l^- = 0$ for all l. If $x_i \leq g_l(x)$ then (5c) and (5a) give $F_{li} \geq 0$, with equality when $x_i = g_l(x)$.

Otherwise, $g_l(x) < 0$ giving $h = 0$ and, by (5a), $F_{li} \geq 0$.

IV. Utilization and fairness

In this section, we discuss some implications of the results in Section III on link utilization and fairness of the equilibrium rates. Theorem 5 shows that XCP equilibrium is constrained max-min fair. It is instructive to compare the XCP equilibrium with the (standard) max-min fair allocation and a class of algorithms proposed in [15].

It is proved in [15] that a (standard) max-min fair rate vector x^* is the unique solution of the same hierarchy of problems $\mathbf{P}_n$ (8)–(12) defined in Section III, except that the feasible set X_0 in (7) is replaced with the superset

$$\overline{X}_0 \;:=\; \{\, x \in \Re_+^N \mid Rx \leq c \,\} \qquad (24)$$

The key feature that results from this much simpler feasible set $\overline{X}_0$ is that the bottleneck links under a max-min fair allocation are all fully utilized. Indeed, a rate vector $x^* \in \overline{X}_0$ is max-min fair if and only if, for every source i, there is a link $l \in L(i)$ in its path such that [14]

1) $y_l(x^*) = c_l$
2) $x_i^* \geq x_j^*$ for all $j \in I(l)$,

From Theorem 5, condition 1 is replaced with the fixed point equation $x_i^* = g_l(y_l(x^*))$ for XCP equilibrium. The simpler condition for max-min fairness has several implications.

First it allows a much simpler proof of max-min fair vector as the unique solution of the problems $\mathbf{P}_n$; see [15]. Second the (centralized) algorithm to compute the max-min fair rate vector (see [15], [14]) is simpler than that in Section III-C for the constrained max-min fair vector. Third, and most importantly, the XCP equilibrium can under-utilize link capacities and deviate by an arbitrarily large factor from the max-min fair allocation, as we illustrate below.

Max-min fairness is generalized in [15] by restricting the feasible set to a (strict) subset of $\overline{X}_0$ in (24). Like XCP, the restriction is specified as additional constraints on source rates x_i and link flows y_l. An example is that, in addition to being in $\overline{X}_0$, a feasible rate vector x must also satisfy

$$x_i \;\leq\; \frac{1}{s^2 c_l}(c_l - y_l)^2 \qquad \forall i, \forall l \in L(i)$$

This is motivated by an explicit design objective of trading off full link utilization for the ability to accommodate random rate fluctuations. If the standard deviation of the rate of source i is sx_i, then it is shown in [15] that the standard deviation of the link flow y_l is less than the spare capacity $c_l - y_l$, so that overshoot is avoided, i.e., $y_l(t) \leq c_l$ for all t in the absence of feedback delay. An alternative additional constraint in [15] is

$$x_i \;\leq\; \frac{f_i}{e_l}(c_l - y_l) \qquad \forall i, \forall l \in L(i)$$

This is again motivated by an explicit design objective: the link parameter e_l controls utilization and source parameter f_i controls fairness, akin to XCP's efficiency and fairness controllers. A distributed algorithm to compute the equilibrium rates is also provided in [15], and its convergence proved. Like XCP, explicit feedback is required: each link l feeds back the spare capacity $c_l - y_l(t)$ to sources that go through this link. Sources adjust their individual rates based on feedback on its path in a way that is distributed, yet avoids overshoot.

We now illustrate the effect of the additional constraint (7) in XCP on link utilization and fairness.

As we explained in the proof of Theorem 3, there are three types of links. The first type are bottlenecks for all the flows that go through that link. All links of this type, such as all $l \in L_1$ in problem $\mathbf{P}_1$, are fully utilized, $y_l = c_l$. The second type are bottlenecks for none of the flows that go through that link. They are underutilized, $y_l < c_l$, because the flow rates going through the link are constrained elsewhere. The third type are bottlenecks for some, but not all, of the flows that go through the link. In contrast to the standard max-min fair allocation, these links are also underutilized, $y_l < c_l$. We can bound the utilization of these partial bottlenecks.

Theorem 7: If $l \in L_1(i)$ for some i then

$$\frac{\alpha}{\gamma\rho_l + \alpha} \;\leq\; \frac{y_l}{c_l} \;\leq\; 1 - \frac{\gamma\sigma_l(\rho_l - \sigma_l)}{\gamma\rho_l + \alpha}$$

Proof: Noting that $\rho_l < 1$ (and that $2(\gamma\rho_l + \alpha) > 0$), removing the last term from the square root in (15) gives the lower bound:

$$\begin{aligned} \frac{y_l}{c_l} &\geq \frac{\alpha + (\gamma+\alpha)\sigma_l + |(\alpha - (\gamma+\alpha)\sigma_l)|}{2(\gamma\rho_l + \alpha)} \\ &\geq \frac{\alpha}{\gamma\rho_l + \alpha} \end{aligned} \qquad (25)$$

where the second inequality is an equality if $\sigma \leq \alpha/(\gamma+\alpha)$.

To derive the upper bound, first note that $\rho_l \geq \sigma_l$ from Lemma 2(2). Since $2(1-\rho_l)(\gamma\rho_l+\alpha) > 0$ and $\Xi \geq 0$, removing the last term from the square root of (16) yields

$$\frac{r^l N_l}{c_l} \leq \frac{1-\sigma_l}{1-\rho_l} - \frac{\gamma\sigma_l(\rho_l-\sigma_l)}{(1-\rho_l)(\gamma\rho_l+\alpha)} \qquad (26)$$

Multiplying both sides by $(1-\rho_l)$ and adding σ_l lead to the upper bound on utilization

$$\frac{y_l}{c_l} \leq 1 - \frac{\gamma\sigma_l(\rho_l-\sigma_l)}{\gamma\rho_l+\alpha} \qquad (27)$$

■

Substituting either $\gamma = 0$ or $\rho_l = \sigma_l$ into either the exact expressions (15) and (16) or the upper and lower bounds (25) and (26) gives full utilization as in the max-min case: $y_l = c_l$ and $r^l = c_l(1-\sigma_l)/(N_l(1-\rho_l))$. This shows that XCP could be made to approach max-min fairness if the bandwidth shuffling were reduced.

On the other hand, link utilization could be arbitrarily low if α and γ had been chosen poorly. With the values suggested in [9] however the utilization is at least 80%. Consider a network of two links. Link 1 has $c_1 = 1$ and carries N_1 flows, while link 2 has $c_2 = 1+\gamma/\alpha$ and carries $N_2 = N_1 + 1$ flows, consisting of all the traffic on link 1 plus one other flow. As $N_1 \to \infty$ we get $\rho_2 \to 1$. This gives $\sigma_2 = \alpha/(\gamma+\alpha)$ in the limit. Thus, both terms in the square root of (15) go to zero, and (25) becomes tight, and $y_l/c_l \to 0$ as $\gamma/\alpha \to \infty$. However, with $\alpha = 0.4$ and $\gamma = 0.1$ [9], (15) gives $y_l/c_l = 0.8$.

Similarly, a given flow may obtain an arbitrarily small proportion of its max-min fair bandwidth for *any* $\alpha > 0$ and $\gamma > 0$. The ratio of the upper bound on XCP bandwidth (26) to the max-min fair bandwidth, $r^{l,mm} = c_l(1-\sigma_l)/(N_l(1-\rho_l))$, is minimized with respect to σ_l when $\rho_l = 2\sigma_l - \sigma_l^2$. Substituting this value into (16) and dividing by $r^{l,mm}$ gives

$$\frac{r^l}{r^{l,mm}} = \frac{C - \frac{D}{1-\sigma_l} + \sqrt{\left[C + \frac{D}{1-\sigma_l}\right]^2 - \frac{E}{1-\sigma_l}}}{2(\gamma(2\sigma_l - \sigma_l^2) + \alpha)} \qquad (28)$$

where

$$\begin{aligned} C &= \Xi_l/(1-\sigma_l) = \gamma\sigma_l(1-\sigma_l) + \alpha, \\ D &= \gamma\sigma_l^2(2-\sigma_l) \\ E &= 4\gamma\sigma_l^2\alpha. \end{aligned}$$

Thus

$$\frac{r^l}{r^{l,mm}} = \frac{C - \frac{D}{1-\sigma_l} + \frac{D}{1-\sigma_l}\sqrt{1 + \frac{(1-\sigma_l)^2}{D^2}\left(C^2 + \frac{2CD-E}{1-\sigma_l}\right)}}{2(\gamma(2\sigma_l - \sigma_l^2) + \alpha)}$$

Applying the identity $\sqrt{1+x} \leq 1 + x/2$, for $x \geq -1$, gives

$$\frac{r^l}{r^{l,mm}} \leq \frac{C + \frac{1-\sigma_l}{2D}C^2 + \frac{2CD-E}{2D}}{2(\gamma\sigma_l(2-\sigma_l) + \alpha)}. \qquad (29)$$

In the limit as $\sigma_l \to 1$, the right hand side tends to 0 for any $\gamma \neq 0$. This demonstrates that, for any non-zero amount of bandwidth shuffling, XCP can be arbitrarily unfair for some topology.

Hence, although the equilibrium of (4) converges to max-min as $\gamma \to 0$, this convergence is not uniform with respect to topology. In other words, given any topology specified by (R, c), we can choose γ sufficiently small so that the resulting allocation is close to max-min fairness. However, for any fixed $\gamma > 0$, such as 0.1 used by XCP, there are topologies in which some source rates can be far away from their max-min allocations.

This behavior can be exhibited by a simple two link network: one link has capacity 1 and carries n^2 flows, while the other carries $n^2 - 1$ of those same flows and has capacity $(n-1)/n$. This network has $\sigma_2 = (n-1)/n$ and $\rho_2 = (n^2-1)/n^2 = 2\sigma_2 - \sigma_2^2$. Hence, $\sigma_2 \to 1$ as $n \to \infty$ and $r^2/r^{2,mm} \to 0$.

These asymptotic results will be illustrated and confirmed by simulation in the following section.

V. Simulation results

This section presents simulation results using the implementation from [9] in NS-2. These results verify the accuracy of our algorithm in Section III-C and confirm our qualitative discussion in Section IV on the utilization and fairness properties of XCP.

We assume that all sources always have packets to send. The topology used for Scenarios 1,2 and 3 is shown in Figure 1 and consists of two links, with $i+j$ sources traversing link L1 and j sources traversing L2.

All links have equal propagation delay of $d_l = 50$ms in both directions and the variable `avg_rrt_` in the XCP implementation (d in the analysis) is fixed to be $4d_l$. The XCP default parameters $\alpha = 0.4$, $\beta = 0.226$ and $\gamma = 0.1$ are used. Although the analysis neglects the "residual" terms, the simulations include them. However, as remarked in Section II-B, they have minimal impact on equilibrium properties. This is confirmed by the good match between theory and simulation as we will see below.

Scenario 1 investigates the utilization of L1 as the number of sources traversing L1 and L2 is changed. In the experiment $i \geq j$, with c_1 =155Mbps and c_2 =100Mbps. The utilization of L1 for a range of i and j is shown in Figure 2. A max-min fair allocation would result in a full utilization of L1 for all i and j combinations. However, as the number of sources bottlenecked at L2 increases, XCP's utilization of L1 decreases.

Since XCP's "residual" terms depend on feedback from upstream nodes, the equilibrium rates depend on the order in which links are traversed. If the direction of flow in this network were reversed, then the utilization would be 0–4% higher than for the case considered and than the theoretical predictions.

Scenario 2 demonstrates that XCP can be arbitrarily unfair for some topology. Let c_1 =155Mbps, $c_2 = c_1(n-1)/n$, $i = n^2 - 1$ and $j = 1$. The ratio of the rate of the source traversing only L1 to the max-min fair rate is plotted in Figure 3. Indeed the unfairness increases with the number of sources in the network, confirming the theory.

Scenario 3 studies XCP with non-standard parameters. It verifies that $y_2/c_2 \to 0$ as $\gamma/\alpha \to \infty$. We set c_2 =200Mbps, $c_1 = c_2(1 + \gamma/\alpha)$, $i = 256$ and $j = 1$. The parameter α is varied from 0.512 to 0.016 and the utilization of L1 as a function of γ/α, as well as the lower bound from (25), are plotted in Figure 4.

Scenario 4 tests the rate allocation algorithm for a more complicated topology as shown in 5. The link one-way delays in ms are $d_1 = 2.5$, $d_2 = 5$, $d_3 = 10$, $d_4 = 10$, $d_5 = 20$ and $d_6 = 40$. The link capacities in Mbps are $c_1 = 10$, $c_3 = 8$, $c_4 = 8$, $c_5 = 7$, $c_6 = 6$ and c_2 is varied in this experiment. The source rates are plotted in Figure 6. There is a good agreement between the predicted and measured rates even though the lower bandwidth delay product makes the fluid flow approximation more questionable.

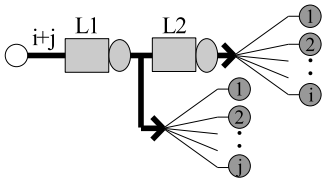

Fig. 1. Topology for Scenarios 1, 2, and 3.

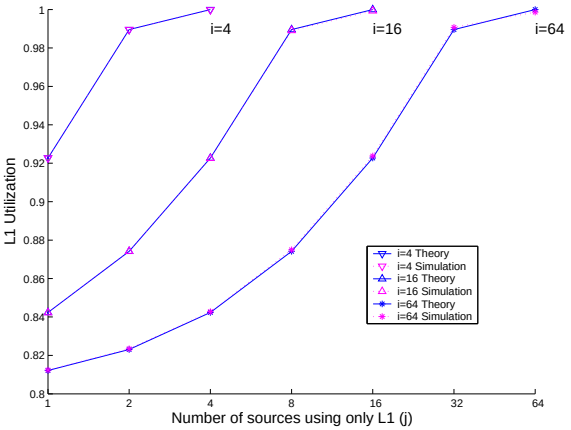

Fig. 2. Scenario 1: utilization.

VI. Conclusion

We have presented a dynamic model of XCP and used it to completely characterize its equilibrium properties. We have shown that XCP clears the queues in equilibrium, and has unique equilibrium rates that solve a constrained max-min fairness problem. The additional constraint under XCP can lead to unfairness for some network topologies. XCP gives a utilization of at least 80%, but a poor choice of α or γ could lead to arbitrarily low utilization. We have provided an algorithm to compute the equilibrium for general networks, and have presented simulation results to illustrate these findings.

VII. Acknowledgment

This work is supported by NSF, ARO, AFOSR, Cisco, and the Caltech Lee Center for Advanced Networking as part of the FAST Project, and supported by the Australian Research Council. We thank Dina Katabi of MIT for helpful discussions.

References

[1] V. Jacobson, "Congestion avoidance and control," *Proceedings of SIGCOMM'88, ACM*, August 1988. An updated version is available via `ftp://ftp.ee.lbl.gov/papers/congavoid.ps.Z`.

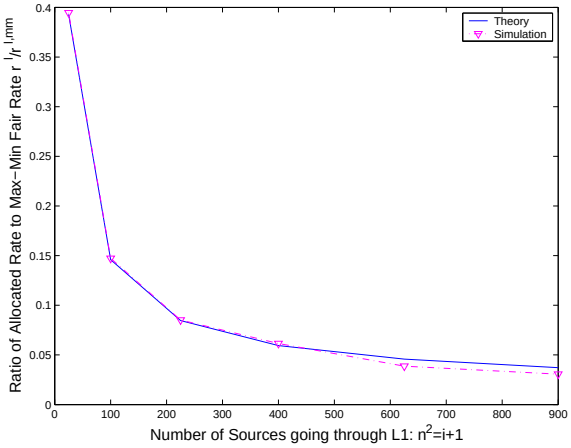

Fig. 3. Scenario 2: unfairness, $r^2/r^{2,mm} \to 0$ as $n \to \infty$.

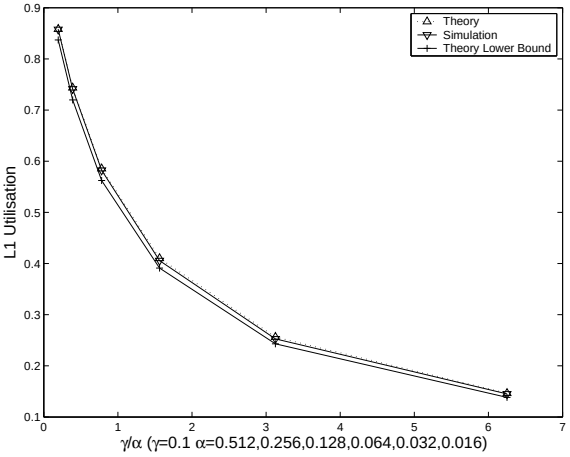

Fig. 4. Scenario 3: utilization of L1 as function of γ/α. With default parameter, $\gamma/\alpha = 0.25$ and utilization = 80%.

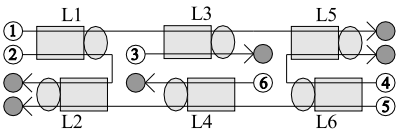

Fig. 5. Scenario 4 topology.

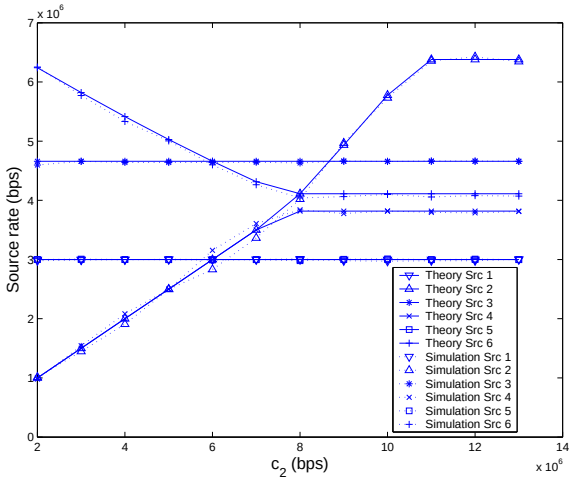

Fig. 6. Scenario 4: throughputs.

[2] C. Hollot, V. Misra, D. Towsley, and W. Gong, "Analysis and design of controllers for AQM routers supporting TCP flows," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 945–959, 2002.

[3] S. H. Low, F. Paganini, J. Wang, and J. C. Doyle, "Linear stability of TCP/RED and a scalable control," *Computer Networks Journal*, vol. 43, no. 5, pp. 633–647, 2003. `http://netlab.caltech.edu`.

[4] C. Casetti, M. Gerla, S. Mascolo, M. Sansadidi, and R. Wang, "TCP Westwood: end-to-end congestion control for wired/wireless networks," *Wireless Networks Journal*, vol. 8, pp. 467–479, 2002.

[5] S. Floyd, "HighSpeed TCP for large congestion windows." Internet draft draft-fbyd-tcp-highspeed-02.txt, work in progress, `http://www.icir.org/floyd/hstcp.html`, February 2003.

[6] C. Jin, D. X. Wei, and S. H. Low, "TCP FAST: motivation, architecture, algorithms, performance," in *Proceedings of IEEE Infocom*, March 2004. `http://netlab.caltech.edu`.

[7] L. Xu, K. Harfoush, and I. Rhee, "Binary increase congestion control for fast long-distance networks," in *Proc. of IEEE Infocom*, 2004.

[8] T. Kelly, "Scalable TCP: Improving performance in highspeed wide area networks." Submitted for publication, `http://www-lce.eng.cam.ac.uk/~ctk21/scalable/`, December 2002.

[9] D. Katabi, M. Handley, and C. Rohrs, "Congestion control for high-bandwidth delay product networks," in *Proc. ACM Sigcomm*, August 2002.

[10] B. Wydrowski and M. Zukerman, "MaxNet: A congestion control architecture for maxmin fairness," *IEEE Communications Letters*, vol. 6, pp. 512–514, November 2002.

[11] B. Wydrowski, L. L. H. Andrew, and M. Zukerman, "MaxNet: A congestion control architecture for scalable networks," *IEEE Communications Letters*, vol. 7, pp. 511–513, October 2003.

[12] S. Low, L. Andrew, and B. Wydrowski, "Understanding XCP: Equilibrium and fairness." `http://netlab.caltech.edu/pub/papers/XCP040505.ps`, 2004.

[13] S. Athuraliya, V. H. Li, S. H. Low, and Q. Yin, "REM: active queue management," *IEEE Network*, vol. 15, pp. 48–53, May/June 2001. Extended version in *Proceedings of ITC17*, Salvador, Brazil, September 2001. `http://netlab.caltech.edu`.

[14] D. Bertsekas and R. Gallager, *Data Networks*. Prentice-Hall Inc., 2nd ed. ed., 1992.

[15] E. M. Gafni and D. P. Bertsekas, "Dynamic control of session input rates in communication networks," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 1009–1016, Jan. 1984.