Path selection and multipath congestion control

Peter Key
Microsoft Research
Cambridge, UK

Laurent Massoulié
Thomspon Technology,
Paris, France

Don Towsley*
University of Massachusetts
Amherst, MA, USA

***Abstract*— In this paper we investigate the potential benefits of coordinated congestion control for multipath data transfers, and contrast with uncoordinated control. For static random path selections, we show the worst-case throughput performance of uncoordinated control behaves as if each user had but a single path (scaling like $\log(\log(N))/\log(N)$ where N is the system size, measured in number of resources). Whereas coordinated control gives a throughput allocation bounded away from zero, improving on both uncoordinated control and on the greedy-least loaded path selection of e.g. Mitzenmacher. We then allow users to change their set of routes and introduce the notion of a Nash equilibrium. We show that with RTT bias (as in TCP Reno), uncoordinated control can lead to inefficient equilibria. With *no* RTT bias, both uncoordinated or coordinated Nash equilibria correspond to desirable welfare maximising states. Moreover, simple path reselection polices that shift to paths with higher net benefit can find these states.**

I. Introduction

Multipath routing architectures have received attention recently [2], [6], [16], [17], [22]. There is considerable interest in combining multipath routing with rate control, e.g. [7], [11], [13]. It can be viewed as an example of cross-layer optimisation [5], [18], where additional benefits are obtained by jointly optimising at the routing (network) and transport layers. Indeed, it is implicitly used in several Peer-to-peer (PTP) applications, in a receiver-driven mode. An early example is Kazaa which allowed users to choose multiple paths, with path selection effectively manual. More recent P2P applications such as Skype use automatic path selection; Skype [21] claims to keep multiple connections open and dynamically chooses the "best" path in terms of latency/quality. Bittorrent [4] maintains 4 active paths with an additional path periodically chosen at random together with a mechanism that retains the best paths (as measured by throughput).

In all the above, users or the end-system's protocol are effectively provided with a large set of potential paths from which they choose a small set with the option of trying to improve upon them. In each case, TCP is used as the transport protocol. Some natural questions arise:

- How does such a mechanism perform relative to one that simply opens and uses all paths? Opening multitudinous TCP connections has systems performance implications, hence there are incentives to keep this overhead small.
- What is effect of RTT bias, if any? TCP Reno has a built-in bias against long RTTs, and we would like to explore the implications.
- And how does it perform relative to a mechanism that uses a *coordinated controller*? By a coordinated controller, we mean one that *actively* balances load across a set of paths, taking into account the states of all paths. A coordinated congestion controller requires a revised transport layer protocol or an application layer solution. In contrast an uncoordinated controller can be thought of as using parallel connections. [1]

The motivating application scenario is of data transfers over TCP, where the transfers are long enough to allow performance benefits for multipath routing. However our analysis applies more generally to situations where there are alternative resources which can help service a demand, and where the demand is serviced using some form of rate control. We assume that the demand is fixed, and each user is attempting to optimise its performance by choosing appropriate paths (resources), where the rate control algorithm is fixed. More precisely, we assume that the rate control is implicitly characterised by a utility maximisation problem [20], where a particular rate control algorithm (eg TCP Reno) is mapped to a particular (user) utility function [8], and that users selfishly seek to choose paths in such a way as to maximise their net utility. A coordinated controller is modelled by a single utility function per user, whose argument is the aggregate rate summed over paths, whereas an uncoordinated controller has a utility function per path and the aggregation is over the utility functions.

* This material is based on work supported by NSF Grant No. CNS-0519922. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

[1] Parallel connections achieve a more limited form of load balancing. For example, two uncoordinated TCP connections, with different throughput rates, in parallel to download a file, such as by pulling from different ends of the file, will cause more of the file to be downloaded on the better path.

We first consider the case where the integer number of paths (resources selected) chosen is fixed at b, and the paths are static, but chosen at random from a set of size N. We look at the worst-case allocation, which is a measure of the fairness of the scheme. In the uncoordinated case, we show that the worst case allocation scales as $\log(\log(N))/\log(N)$, with increasing b only improving the constant in the scaling. In contrast, in the coordinated case where we can rebalance the load across resources, provided $b > 1$, the worst-case allocation is bounded away from zero. This demonstrates that

1) coordinated improves significantly on uncoordinated in the static case;
2) coordinated improves on the greedy least-loaded resource selection, as in Mitzenmacher [19], where the least-loaded selection of b resources scales as $1/\log(\log(N))$ for $b > 1$.

Effectively, the coordinated selection is able to shift the load amongst the resources, and with a minimal choice of b able to utilise the resources *as if* a global load balance was being performed.

We then allow users to change the set of routes they use, and introduce the natural game-theoretic notion of a Nash equilibrium in this context, where users seek to selfishly maximise their own net utilities. We then find qualitatively different behaviour according to whether the rate controller has a RTT bias or not. If the controller has no RTT bias (unlike TCP Reno, for example), then in the uncoordinated case we find that the Nash equilibria correspond to desirable welfare maximising states, which implies we have a Pareto-efficient solution. In contrast, if there is an RTT bias (as in TCP Reno), then the Nash equilibria can be inefficient, and we give an example where the achieved rate is half of what could be achieved. For the coordinated controller, of necessity there can be no RTT bias, and Nash equilibria coincide with welfare-maximising social optima. Moreover, we show for both coordinated and uncoordinated control *with no RTT bias*, that simple path selection policies which combine random path resampling with moving to paths with higher net benefit lead to welfare maximising equilibria, and do as well as if the entire path choice was available to each user.

In summary, we shall provide some partial answers to our initial questions.

- In a large system, provided we re-select randomly from the set of paths and shift between paths with higher net benefit, we can use a small number of paths to choose from and do as well as if we were fully using all the paths
- There is a loss of efficiency with RTT bias
- Coordinated control has better fairness properties than uncoordinated in the static case. When combined with path reselection, uncoordinated control only does as well as a coordinated control if there is no RTT bias in the controllers.

The last two points suggest good design choices for new multipath rate controllers are coordinated controllers or uncoordinated controllers with the RTT bias removed.

We now describe the modelling framework.

II. Modelling Framework

A. Model

We assume a set of user classes, indexed by $s \in \mathcal{S}$. Network paths are indexed by $r \in \mathcal{R}$. Users of class s can use any path from subset $\mathcal{R}(s)$ of $\mathcal{R}$. Without loss of generality we may assume these sets are disjoint. Network capacities or feedback signals (such as loss, packet marking or delay) are captured by some convex non-decreasing penalty function $\Gamma : \mathbb{R}_+^{\mathcal{R}} \to \mathbb{R}_+ \cup \{+\infty\}$ - see [10] for examples. Typically Γ is the sum of penalty functions associated with each resource type. We can also interpret the penalty functions as "costs" and their derivatives as "prices", and we shall make use of the notation

$$p_r = \partial_r \Gamma(\Lambda).$$

B. Uncoordinated Congestion Control

We assume that class s-users try to maximise their throughput. We further assume that each class s-user is restricted to use the same number, b, of connections, that can be along any routes r from set $\mathcal{R}(s)$.

We further assume that the rate a user obtains on a connection along a given route is achieved by some default congestion control mechanism (e.g. TCP), that implicitly performs some utility maximisation, the utility of a single user sending rate λ_r through route r being $U_r(\lambda_r)$. For tractability, we assume that U_r is a strictly concave increasing function that is continuously differentiable on $(0, \infty)$. Finally, we assume that users' criterion for optimality is their achieved rate.

If there are N_s' class s-users, the total number of connections they make is $N_s := bN_s'$. Denoting by N_r for $r \in \mathcal{R}(s)$ the total number of connections made by class s users along route r, gives the following constraint:

$$\sum_{r \in \mathcal{R}(s)} N_r = N_s, \quad s \in \mathcal{S}. \tag{1}$$

The outcome of congestion control for *given* numbers N_r of connections along each route r, is defined to be the solution of the welfare maximisation problem

$$\text{Maximise} \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{R}(s)} N_r U_r(\Lambda_r/N_r) - \Gamma(\Lambda). \tag{2}$$

over $\Lambda_r \geq 0$ where $\Lambda = \{\Lambda_r\}$ denotes the vector of aggregate rates. Note that the utility function can depend upon the route taken.

The function being optimised in (2) is a strictly concave function, optimised over a convex feasible region; hence the problem is Strong Lagrangean and the unique maximum is attained. Moreover, the function is also strictly concave over (Λ, N), provided $N > 0$, where $N_r U_r(\Lambda_r/N_r)$ for $N_r > 0$ is the *perspective* (page 89, [3]) of U_r; hence we can consider the optimisation (2) subject to (1) for (Λ, N) over $\Lambda \geq 0, N > 0$ to look at the optimal choice of paths and rates.

C. Coordinated Congestion Control

Assume that class s-users can use concurrently paths from a collection c, where $c \subset \mathcal{R}(s)$, and denote by $\mathcal{C}(s)$ the family of all such path collections that are allowed. For definiteness, think of $\mathcal{C}(s)$ as the collection of all subsets of $\mathcal{R}(s)$ of size b. Denote by N_c the number of users with associated set of connections equal to c. When the number of class s users equals N_s, one thus has the constraint

$$\sum_{c \in \mathcal{C}(s)} N_c = N_s, \quad s \in \mathcal{S}. \tag{3}$$

In contrast to the uncoordinated case, we associate a single utility function $U_s(\cdot)$ with a class s user, assumed strictly concave, increasing, and continuously differentiable on $(0, \infty)$. We can then assume that the allocation to a class s-user with connection set c is $\sum_{r \in c} \Lambda_{c,r}/N_c$, where the quantities $\Lambda_{c,r}$ solve

$$\text{Maximise} \sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}(s)} N_c U_s \left(\frac{\sum_{r \in c} \Lambda_{c,r}}{N_c} \right) - \Gamma(\Lambda) \tag{4}$$

over $(\Lambda_{c,r} \geq 0$, where the vector $\Lambda = (\Lambda_r)$ in the argument of the penalty function Γ is defined as

$$\Lambda_r := \sum_{c: r \in c} \Lambda_{c,r}. \tag{5}$$

The joint optimisation (4) over $\Lambda \geq 0, N > 0$ subject to (5) is also Strong Lagrangean. We shall see in Section V, that the optimal rates Λ_r for this joint optimisation actually solve the following welfare maximisation

$$\text{Maximize} \sum_{s} N_s U_s \left(\frac{\sum_{r \in \mathcal{R}(s)} \Lambda_r}{N_s} \right) - \Gamma(\Lambda) \tag{6}$$

$$\text{over } \Lambda_r \geq 0, \quad r \in \mathcal{R}. \tag{7}$$

This problem is strong Lagrangean, and its solution is characterized by the Kuhn-Tucker conditions

$$U_s' \left(\frac{\sum_{r \in \mathcal{R}(s)} \Lambda_s}{N_s} \right) \leq \partial_r \Gamma(\Lambda), \tag{8}$$

$$U_s' \left(\frac{\sum_{r \in \mathcal{R}(s)} \Lambda_s}{N_s} \right) < \partial_r \Gamma(\Lambda) \Rightarrow \Lambda_r = 0. \tag{9}$$

We note that there are distributed rate control algorithms for all of the above optimisation problems, e.g. [11].

III. Static, random route selections

In this section we focus on the following scenario. There are N resources with unit capacity, and the penalty function associated with each resource is the step function,

$$\Gamma_r(\Lambda_r) = \begin{cases} 0 & \text{if } \Lambda_r \leq 1 \\ \infty & \text{otherwise.} \end{cases} \tag{10}$$

To provide a concrete interpretation, the resources can be interpreted as servers, or as relay or access nodes. There are aN users. Each user selects b resources at random from the N available, where b is an integer larger than 1 (the same resource may be sampled several times). We shall look at the worst case rate allocation of users under two distinct bandwidth sharing scenarios. In the first scenario, there is no coordination between the distinct b connections of each user. Thus, if one connection uses a resource handling X connections overall, it is straightforward to show that the connection achieves a rate allocation of exactly $1/X$. In the second scenario, each user implements coordinated multipath congestion control.

The worst-case allocation, is a fairness measure. However, for our scenario it is straightforward to show [14] that the more "unfair" the allocation, the greater the expected time to download a unit of data, and that a coordinated allocation minimises such a performance measure.

A. Uncoordinated congestion control

We shall denote by λ_i the total rate that user i obtains from all its connections. The main result is the following

Theorem 3.1: For fixed parameters a and b, then for any $\epsilon > 0$, one has the following

$$\lim_{N \to \infty} \mathbf{P} \left(\min_{i=1,\ldots,aN} \lambda_i \leq (b^2 + \epsilon) \frac{\log(\log(N))}{\log(N)} \right) = 1. \tag{11}$$

In words, the worst case allocation in this scenario decreases like $\log(\log(N))/\log(N)$. This is to be compared with the worst case allocation that one gets if $b = 1$, that is if a single path is used: from classical balls and bins models [19], this also decreases in $\log(\log(N))/\log(N)$ as N increases.

The proof relies on the following result:

Lemma 3.1: Let the constant $b > 0$ be as above. Given some constant $\alpha > 0$, when αN balls are thrown at random in N bins, then for any $\epsilon > 0$, $\epsilon < 1/2b$, with high probability, the number of bins which receive at least

$$M := \left(\frac{1}{b} - 2\epsilon \right) \frac{\log(N)}{\log(\log(N))} \tag{12}$$

balls is larger than $N^{1+\epsilon-1/b}$.

Proof: Let $\beta = \alpha/2$, and let X be a Poisson random variable with mean βN. Then if one throws

X balls at random into N bins, by a standard property of Poisson random variables, the numbers of balls in each bin are i.i.d. random variables, admitting a Poisson distribution with parameter β. Furthermore, with high probability, $X \leq \alpha N$, so it is enough to show that the property of the Lemma holds in the case where the occupancy numbers of bins are i.i.d, Poisson with parameter β.

In this context, the number of bins that receive at least M balls has a Binomial distribution, with parameters (N, q) where

$$q = \mathbf{P}(\text{Poisson}(\beta) \geq M) \geq e^{-\beta}\frac{\beta^M}{M!}.$$

As follows from standard Chernoff bounds (see eg [1], appendix), Binomial random variables admit tighter Chernoff bounds than Poisson variables with the same mean. Hence, if one can show that the mean Nq of this Binomial random variable satisfies

$$\lim_{N\to\infty} \frac{Nq}{N^{1+\epsilon-1/b}} = +\infty,$$

the result of the Lemma will follow. However, the logarithm of the left-hand side of the above satisfies, appealing to Stirling's formula, and neglecting lower order terms:

$$\begin{aligned}
\log\left(\tfrac{Nq}{N^{1+\epsilon-1/b}}\right) &\geq -\beta + M\log(\beta) - \log(M!) \\
&\quad +(1/b-\epsilon)\log(N) \\
&\sim -\beta + M\log(\beta) - \tfrac{1}{2}\log(2\pi M) \\
&\quad -M\log(M/e) + (1/b-\epsilon)\log(N) \\
&\sim -(1/b-2\epsilon)\log(N) \\
&\quad +(1/b-\epsilon)\log(N) \\
&= \epsilon\log(N),
\end{aligned}$$

where we have used the expression (12) of M. This establishes the result. ∎

Proof: (of Theorem 3.1) Assume that the resource selection of hosts is broken into two phases. First, one half (that is, γN users) make their individual selections. By the above lemma, once this is done, with high probability there are at least $N^{1+\epsilon-1/b}$ resources selected by at least M users, at the end of this phase. In the second phase, the remaining γN users get to select their resources. Each random selection of these users therefore has a probability of at least $N^{1+\epsilon-1/b}/N = N^{\epsilon-1/b}$ of being to a resource with at least M users. Thus the probability that such a user makes selections only to resources with at least M users is at least $N^{b\epsilon-1}$.

Therefore, the total number of phase two users selecting only such congested resources is larger than a Binomial random variable with parameters $(\gamma N, N^{b\epsilon-1})$. As its mean $\gamma N^{b\epsilon}$ goes to $+\infty$ as $N \to \infty$, there is at least such a user with high probability. Its total rate allocation will then be at most

$$\lambda = \frac{b}{M} = \frac{b}{1/b-2\epsilon}\frac{\log(\log(N))}{\log(N)}.$$

For any $\epsilon' > 0$, by taking $\epsilon > 0$ small enough, the first fraction in the above is indeed less than $b^2 + \epsilon'$, which completes the proof. ∎

B. Coordinated congestion control

Here we assume as before that there are aN users, each selecting b resources at random, from a collection of N available resources. We shall denote by λ_{ij} the rate that user i obtains from resource j, and let A_{ij} equal 1 if user i accesses resource j, and equal 0 otherwise.

In contrast with the previous situation, we now assume that the rates λ_{ij} are chosen to maximize:

$$\sum_{i=1}^{aN} U\left(\sum_{j=1}^{N} A_{ij}\lambda_{ij}\right)$$

under the constraints:

$$\lambda_{ij} \geq 0, \quad \sum_{k=1}^{aN} A_{kj}\lambda_{kj} \leq 1, \quad i \leq aN, j \leq N.$$

In the above, U is a strictly concave, increasing utility function, and an insensitivity result shows that the allocation is independent of the particular utility function chosen, whose proof we omit for brevity.

This in turn implies the following characterisation of the optimal rates (λ_i^*) as the so-called max-min fair allocations, whose proof we also omit:

Lemma 3.2: Let (λ_i^*) be the optimal user rates solving the above optimisation problem for some (and hence for all) strictly concave, increasing utility function U. Denote by $x_1 < x_2 < \cdots < x_m$ the distinct values of the λ_i^*, ranked in increasing order. Let I_k denote the set of indices i such that $\lambda_i^* = x_k$.

Then for any other feasible allocation (λ_i), necessarily

$$\min_{i\in I_1}(\lambda_i) \leq x_1.$$

If there is equality in the above, , $\lambda_i \equiv x_1$ on I_1.
We can now establish the following:

Theorem 3.2: Assume there are N resources, and aN users each connecting to b resources selected at random. Denote by λ_i^* the optimal allocations that result. Then there exists $x > 0$, that depends only on a and b, such that:

$$\lim_{N\to\infty} \mathbf{P}\left(\min_i \lambda_i^* \geq x\right) = 1. \tag{13}$$

A sufficient condition for this evaluation to be valid is that $x < \min(1/a, b-1)$, and furthermore:

$$\forall u \in (0, a], ah(u/a) + h(ux) + bu\log(ux) < 0, \tag{14}$$

where $h(x) := -x\log(x) - (1-x)\log(1-x)$ is the classical entropy function.

That is to say, the worst case allocation is bounded away from 0 as N tends to infinity. This should be compared to the result quoted by Mitzenmacher et al. [19], which says that if users arrive in some random order, and choose among their b candidate resources a single one, then the worst case rate scales like $1/\log(\log(N))$. Thus we achieve better use of resources by actively balancing load among several available resources.

Proof: By Hall's theorem, there exists a feasible allocation (λ_i) to users such that $\min_i \lambda_i \geq x$ if and only if, for any set I of user indices, one has:

$$x|I| \leq |\{j : A_{ij} = 1 \text{ for some } i \in I\}|. \qquad (15)$$

By Lemma 3.2, if there exists such an allocation, then necessarily the utility maximising allocation (λ_i^*) must also be such that $\lambda_i^* \geq x$ for all i. It thus remains to prove that for some suitable $x > 0$, with high probability Condition (15) is met for all non-empty subsets $I \subset \{1, \ldots, aN\}$.

For any $k \in \{1, \ldots, aN\}$, let $r_k := \lceil kx \rceil - 1$ be the smallest integer strictly less than kx. Denote by R_i the (random) set of resources that user i tries to connect to. Then the probability that the desired property fails to hold reads:

$$\mathbf{P}(\exists I \subset \{1, \ldots, aN\}, \exists J \subset \{1, \ldots, N\}, \text{so that:}\\ |J| \leq r_{|I|} \text{ and } \cup_{i \in I} R_i \subset J).$$

Note that, for a particular user set I of size k, and a particular resource set J of size r, the probability that all the bk random resource selections made by all users i, $i \in I$, fall into set J, equals $(r/N)^{bk}$. Thus, by the union bound, the above probability is not larger than

$$\sum_{k=1}^{aN} \binom{aN}{k} \binom{N}{r_k} \left(\frac{r_k}{N}\right)^{bk}. \qquad (16)$$

Those terms with $r_k = 0$ are null, and can thus be ignored. Under the condition that $xa < 1$, the second binomial coefficient is non-zero for all k in the summation range.

Thus, by Stirling's formula, the k-th term in this sum is not larger than a constant times $\exp(A(k))$, where:

$$A(k) := aNh(k/(aN)) + Nh(r_k/N) + bk\log(r_k/N).$$

The exponent $A(k)$ also reads:

$$\begin{aligned} A(k) = \quad & \log(N)\,[k + r_k - bk] \\ & -k\log(k) - r_k\log(r_k) + bk\log(r_k) \\ & +(aN-k)\log(1 + k/(aN-k)) \\ & +(N-r_k)\log(1 + r_k/(N-r_k)). \end{aligned}$$

Fix some $\delta \in (0, 1/2)$. Then for $k \in \{1, \ldots, N^\delta\}$, the exponent $A(k)$ is not larger than $(1-2\delta)(k + r_k - bk)\log(N)$, which is less than $(1-2\delta)(1 + x - b)k\log(N)$. Fix now some $\epsilon > 0$. In the range $k \in \{N^\delta, \epsilon N\}$, the exponent $A(k)$ is not larger than $k[(1+x-b)\log(N/k) + C]$, for some constant C. This is not larger than $k[(1+x-b)\log(1/\epsilon) + C]$. Thus, for sufficiently small ϵ, the factor of k in this expression is strictly negative (recall the assumption that $1+x-b < 0$). Finally, in the range $k \in \{\epsilon N, aN\}$, we have

$$A(k) \leq N \sup_{u \in [\epsilon, a]} \left[ah(u/a) + h(xu) + b\log(xu)\right].$$

Provided the supremum in this expression is strictly negative, the sum (16) is, up to a constant factor, not larger than:

$$\sum_{k=1}^{N^\delta} N^{(1-2\delta)(1+x-b)k} + \sum_{k=N^\delta}^{\epsilon N} e^{-kC'} + \sum_{k=\epsilon N}^{aN} e^{-C''N},$$

where C', C'' are positive constants. It clearly follows that the sum (16) goes to zero as N tends to infinity.

It now remains to establish that one can indeed select $x > 0$ small enough such that Condition (14) holds. Argue by contradiction, assuming that for all $x > 0$, there exists $u > 0$ such that:

$$ah(u/a) + h(ux) + bu\log(ux) > 0. \qquad (17)$$

Thus necessarily,

$$u < \frac{ah(u/a) + h(ux) + bu\log(u)}{-b\log(x)}.$$

The numerator in the right-hand side is bounded from above, uniformly in $u \in [0, a]$. This shows that, for small x, u must be of order at most $1/|\log(x)|$ and hence go to zero with x. The left-hand side of (17) reads, for small x, and small u:

$$\begin{aligned} & -u\log(u/a) - (a-u)\log(1-u/a) - ux\log(ux) \\ & -(1-ux)\log(1-ux) + bu\log(ux) \\ & = -u\log(u)[1 + x - b] + O(u) + bu\log(x). \end{aligned}$$

The last term in the above is negative for $x < 1$; for $x < b - 1$, and small enough u, the sum of the first two terms in the last display is also negative. This shows that (17) cannot hold: for small enough x, there exists $u > 0$ such that it fails. This concludes the proof. ■

IV. Nash equilibria for throughput-maximising users

In this section we assume that users can choose the set of routes that they use. We characterise equilibrium allocations, assuming users greedily search for throughput-optimal routes. We show that the same equilibria arise with coordinated congestion control, and uncoordinated unbiased congestion control. Moreover, these equilibria achieve welfare maximisation. In contrast, we exhibit specific network topologies where RTT-biased uncoordinated congestion control yields different, inefficient equilibria. We shall use the models and notation of Section II.

A. Uncoordinated, unbiased congestion control

Under the assumptions of Section II-B we introduce the following notion of Nash equilibrium:

Definition 4.1: The collection of per route connection numbers N_r is a Nash equilibrium for selfish throughput maximisation if it satisfies (1), and furthermore, the allocations (2) are such that for all $s \in \mathcal{S}$, all $r \in \mathcal{R}(s)$, if $N_r > 0$, then

$$\frac{\Lambda_r}{N_r} = \max_{r' \in \mathcal{R}(s)} \frac{\Lambda_{r'}}{N_{r'}}. \tag{18}$$

$\Diamond$

The intuitive justification for this definition is as follows: any class s-user would maintain a connection along route r only if it cannot find an alternative route r' along which the default congestion control mechanism would allocate a larger rate.

We then have the following result:

Proposition 4.1: Assume that for each $s \in \mathcal{S}$, there is a strictly concave, increasing class utility function U_s such that $U_r \equiv U_s$ for all $r \in \mathcal{R}(s)$. Then for a Nash equilibrium (N_r), the corresponding rate allocations (Λ_r) solve the general optimisation problem (6-7).

Proof: Let $p_r := \partial_r \Gamma(\Lambda)$. Then for each $r \in \mathcal{R}(s)$ such that $N_r > 0$, it holds, by monotonicity of $U_r' \equiv U_s'$, that

$$p_r = \min_{r' \in \mathcal{R}(s)} p_{r'}.$$

These are precisely the Kuhn-Tucker conditions that characterize maxima of the optimisation problem (6-7). ■

To summarise: if i) the utility functions of the default congestion control mechanism are path-independent, and ii) users agree to concurrently use a fixed number b of paths, and iii) they manage to find throughput-optimal paths, that is they achieve a Nash equilibrium, then at the macroscopic level, the per-class allocations solve the coordinated optimisation problem (6-7).

B. Uncoordinated, biased congestion control

It is well known that the bandwidth shares achieved by TCP Reno are affected by the path round trip time. To illustrate the possible consequences, we follow [8] and assume that TCP implicitly maximises the sum of utilities of all current connections, and the utility of sending at rate λ along a path r with round-trip time τ_r is

$$U_r(\lambda) = -\frac{1}{\tau_r^2}\frac{1}{\lambda}.$$

Consider now the network example of Figure 1. It has long fat links, with associated round trip time T and capacity C, and short thin links, with round trip time τ and capacity c, with $T > \tau$ and $c < C$. Assume class a users need to transfer data from node a to node a', and can use either a "long fat" route via c', b', or a "short thin" route via b, c. Similarly, class b (respectively, c) users need to transfer data from node b to node b' (respectively, from node c to node c') and can take either a route with two short and one long link, or with two long and one short link.

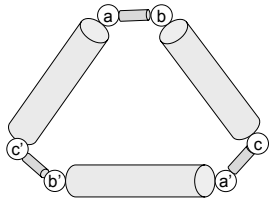

Fig. 1. Network with alternation of fat and long with short and thin links.

Let us now show that bad Nash equilibria can arise for this particular network, given the TCP utility functions described above, and for particular choices of link RTTs. Consider in particular the symmetric case where the numbers of class a, b and c users are all equal to some common number N', all of which use the same number of connections b, and let $N := bN'$. We now demonstrate that the state where all connections are via the (short-long-short) routes is a Nash equilibrium.

In such a state, the number of connections via every short (respectively, long) link is equal to $2N$ (respectively, N). The total round trip time of the (s-l-s) routes is $T + 2\tau$. In order to utilise perfectly the thin links, the corresponding Lagrange multipliers p must satisfy:

$$2p = U'_{s-l-s}(c/2N) = \left[(T + 2\tau)\frac{c}{2N}\right]^{-2},$$

thereby ensuring that each connection achieves a total rate of $c/(2N)$. Consider now the rate that would be achieved on a (long-short-long) route, whose round-trip time is $2T+\tau$, and whose aggregate Lagrange multiplier is p rather than $2p$. The corresponding rate would thus be:

$$\lambda = \frac{1}{(2T + \tau)\sqrt{p}} = \sqrt{2}\frac{T + 2\tau}{2T + \tau}\frac{c}{2N}.$$

Thus, provided the link round trip times τ, T satisfy

$$\sqrt{2}\frac{T + 2\tau}{2T + \tau} < 1,$$

then the state where all connections are of (s-l-s) type is indeed a Nash equilibrium. Note that the total throughput achieved is half what could be achieved using the (l-s-l) paths instead.

C. Coordinated congestion control

For coordinated control, we use the model of section II-C and introduce the following notion of a Nash equilibrium:

Definition 4.2: The non-negative variables N_c, $c \in \mathcal{C}(s)$, $s \in \mathcal{S}$, are a Nash equilibrium for the coordinated congestion control allocation if they satisfy the constraints (3), and moreover, for all $s \in \mathcal{S}$, all $c \in \mathcal{C}(s)$, if $N_c > 0$, then the corresponding coordinated rate allocations satisfy

$$\frac{\sum_{r\in c}\Lambda_{c,r}}{N_c} = \max_{c'\in\mathcal{C}(s)} \frac{\sum_{r\in c'}\Lambda_{c',r}}{N_{c'}}. \qquad (19)$$

$\Diamond$

We then have the following:

Proposition 4.2: At a Nash equilibrium as in Definition 4.2, the path allocations Λ_r solve the welfare maximisation problem (6-7).

Proof: Let $p_r := \partial_r\Gamma(\Lambda)$. Then the allocations of users of type s with connection set c read:

$$\frac{\sum_{r\in c}\Lambda_{c,r}}{N_c} = {U'_s}^{-1}(\min_{r\in c} p_r).$$

Thus the only routes r that type s users utilise at a Nash equilibrium are such that $p_r = \min_{r'\in\mathcal{R}(s)}(p_{r'})$, and all type s-users obtain a global rate equal to ${U'_s}^{-1}(\min_{r\in\mathcal{R}(s)} p_r)$. These are precisely the Kuhn-Tucker optimality conditions for the coordinated welfare optimisation problem (6-7). ■

V. Dynamic route selection

In this section we look at deterministic differential equation models of joint rate adaptation and route selection. We consider first the case of coordinated congestion control and then the case of uncoordinated unbiased congestion control. In both cases, the route selection procedure works as follows: first, a user with a current route set c is proposed a new route set c' at some fixed rate $A_{cc'}$. Then, the new route set is accepted under the condition that the *net benefit* that the user retrieves from the new route set is higher than that of the current route set.

We show for both cases that this procedure eventually leads to a welfare maximising equilibrium.

A. Model

We use the model of Section II-C, where now the number of class s users, N_s, are subdivided according to the set of routes they are currently using, N_c denoting the number of class s-users concurrently using all routes in c, $c \subset \mathcal{R}(s)$. Class s users currently using the set c of routes will at the instants of a Poisson process with intensity $A_{cc'}$ consider replacing their route set c by route set c'. We shall restrict the feasible subsets c of routes that class s users may use by setting to zero some of the $A_{cc'}$ rates, and assume that the feasible route sets have common cardinality b, e.g. $b = 2$. Finally, assume that for each class s, any $r \in \mathcal{R}(s)$, any given set $c \in \mathcal{C}(s)$, there is some c' such that $r \in c'$ and $A_{cc'}$ is positive.

We denote by λ_c the data rate obtained by users streaming along routes $r \in c$. This is the sum of the rates $\lambda_{c,r}$ over $r \in c$, where $\lambda_{c,r}$ is the sending rate along route r:

$$\lambda_c = \sum_{r\in c}\lambda_{c,r}.$$

and related to the aggregate rate Λ_r by

$$\Lambda_r = \sum_{c:r\in c} N_c\lambda_{c,r}, \quad r \in \mathcal{R}.$$

B. Coordinated congestion control

We assume the following form of rate adaptation (see [9]):

$$\frac{d}{dt}\Lambda_{c,r} = N_c\kappa_{c,r}\left[U'_{s(c)}(\Lambda_c/N_c) - \partial_r\Gamma(\Lambda)\right] + \mu_{c,r}, \qquad (20)$$

where the term $\mu_{c,r}$ is non-negative and such that $\mu_{c,r}\lambda_{c,r} \equiv 0$, and is meant to ensure non-negativity of $\lambda_{c,r}$, and $\kappa_{c,r}$ is a positive gain parameter.

We denote the net benefit per unit time for type s users streaming along routes r in some set c as B_c, given by

$$B_c = U_s(\lambda_c) - \sum_{r\in c}\lambda_{c,r}U'_s(\lambda_c).$$

We now make the following assumption. A type s user will swap from route set c to route set c', at an instant where this change is proposed, only if the net benefit $B_{c'}$ exceeds B_c. Note that it may be delicate to do this, and schemes may be needed to actually evaluate such net benefits along the alternative path set c'.

We would thus have a change from $N = \{N_c\}$ to $N + e_{c'} - e_c$ at a rate

$$N_cA_{cc'}\phi(B_{c'} - B_c),$$

where ϕ is a Lipschitz continuous function, taking values in $[0, 1]$, equal to 0 on $\mathbb{R}_-$, and positive on $(0, \infty)$. For definiteness, we shall take

$$\phi(x) = \max(0, \min(x, 1)).$$

Assuming large populations of users, we no longer consider the stochastic system but the deterministic evolution defined by the drift vector field, ie

$$\frac{d}{dt}N_c = \sum_{c'} N_{c'}A_{c'c}\phi(B_c - B_{c'}) - \sum_{c'} N_cA_{cc'}\phi(B_{c'} - B_c). \qquad (21)$$

We now show the following

Proposition 5.1: Assume that the utility functions U_s and the penalty function Γ are continuously differentiable on their domain, that the former are strictly concave increasing, and the latter convex increasing. Assume further that $U'_s(x) \to 0$ as $x \to \infty$. Then any

absolutely continuous solution $(N_c, \lambda_{c,r})$ to the system of ODE's (20-21) converges to the set of maximisers of the welfare function

$$\mathcal{W}(\Lambda, N) := \sum_{s\in\mathcal{S}} \sum_{c\subset\mathcal{R}(s)} N_c U_s(\Lambda_c/N_c) - \Gamma(\Lambda) \quad (22)$$

where $\Lambda_c = N_c\lambda_c$, under the constraints (3). The corresponding equilibrium rates (Λ_r) are solutions of the coordinated welfare maximisation problem (6–7).

Before establishing the proof of this proposition, we provide an interpretation of the net benefit maximisation rule:

Remark 5.1: For any strictly concave, continuously differentiable function U, the corresponding net benefit function $B(x) := U(x) - xU'(x)$ is strictly increasing, as can be seem from writing $B'(x) = -xU''(x)$. Thus, the net benefit maximisation strategy corresponds to a *rate* maximisation strategy.

Proof: Let us first establish that the function $\mathcal{W}$ as defined in (22) increases with time. For almost every t, we have:

$$\frac{d}{dt}\mathcal{W} = \sum_{c\subset\mathcal{R}} \sum_{r\in c} \left(\frac{d}{dt}\Lambda_{c,r}\right) \frac{\partial \mathcal{W}}{\partial \Lambda_{cr}} + \sum_{c\subset\mathcal{R}} \frac{d}{dt} N_c \frac{\partial \mathcal{W}}{\partial N_c} \quad (23)$$

$$= \sum_{c\subset\mathcal{R}} \sum_{r\in c:\lambda_{c,r}>0} \kappa_{c,r} N_c \left[U'_{s(c)}(\lambda_c) - \partial_r\Gamma(\Lambda)\right]^2 \quad (24)$$

$$+ \sum_{c,c'\subset\mathcal{R}} A_{cc'} N_c \phi(B_{c'} - B_c)\left[B_{c'} - B_c\right].$$

Indeed, identity (23) holds by absolute continuity of the functions $t \to \lambda_{c,r}(t)$ and the fact that the welfare function is continuously differentiable. The expression (24) holds because at points t where $\lambda_{c,r}(t) = 0$ and the function $t \to \lambda_{c,r}(t)$ is differentiable, by its non-negativity this derivative must equal zero. Also, to establish (24) we have used the fact that $B_c = \partial\mathcal{W}/\partial N_c$.

Since each term in (24) is non-negative, welfare increases with time as claimed.

We now characterize the limiting points of these dynamics.

Lasalle's invariance theorem (see e.,g. Khalil [15], p.128, Theorem 4.4) ensures that solutions of these ODE's converge to the set of points for which the expression (24) equals zero, provided that the trajectories are bounded. However, boundedness holds trivially for the N-components because the total N_s remains constant, while it holds for the λ-component because of our assumption that $\lim_{x\to\infty} U'_s(x) = 0$.

Thus, solutions of these ODE's converge to the set of points such that for all c such that $N_c > 0$, all $r \in c$, either $\lambda_{c,r} = 0$, or

$$U'_{s(c)}(\lambda_c) = \partial_r\Gamma(\Lambda).$$

This implies that for all $c \subset \mathcal{R}(s)$ such that $N_c > 0$, all $r \in c$ such that $\lambda_{c,r} > 0$, one has:

$$\partial_r\Gamma(\Lambda) \equiv U'_{s(c)}(\lambda_c).$$

Consider now the set of values $p_r = \partial_r\Gamma(\Lambda)$, $r \in \mathcal{R}(s)$, and let r_0 be such that p_{r_0} achieves the minimum of such values. Let $c \subset \mathcal{R}(s)$ be such that, at a candidate equilibrium point, $N_c > 0$. consider the following two cases.

Case 1: $r_0 \in c$. Then necessarily, either $\lambda_c = 0$, and $p_{r_0} \geq U'_s(0)$, or $\lambda_c > 0$, and $U'_s(\lambda_c) = p_{r_0}$.

Case 2: $r_0 \notin c$. However, by assumption there exists $c' \subset \mathcal{R}(s)$ such that $A_{cc'} > 0$ and $r_0 \in c'$. Necessarily, for $N_c > 0$ to hold, one must have

$$B_c \geq B_{c'} = U_s(\lambda_{c'}) - \lambda_{c'}U'_s(\lambda_{c'}).$$

Now in view of Remark 5.1, necessarily $\lambda_c \geq \lambda_{c'}$. In turn, this yields that

$$\min_{r\in c} p_r = U'^{-1}_s(\lambda_c) \leq U'^{-1}_s(\lambda'_c) = p_{r_0}.$$

By our choice of r_0, this implies that $\lambda_c = \lambda_{c'}$, and $p_r = p_{r_0}$ for all $r \in c$ such that $\lambda_{c,r} > 0$.

Consider now the optimisation problem (6),(7) with optimality conditions (9).

From the previous discussion, any point such that $(d/dt)\mathcal{W} = 0$ is such that the corresponding rates

$$\Lambda_r = \sum_{c\subset\mathcal{R}(s)} N_c\lambda_{c,r}$$

solve the above optimization problem. ■

In this sense, random route resampling, coupled with route selection based on net benefit (or by Remark 5.1, based on achieved rate) provides global allocations to user classes that coincide with those that would arise if users were allowed to use coordinated congestion control over the full set of available routes $\mathcal{R}(s)$ simultaneously.

C. Uncoordinated congestion control

We assume the following form of rate adaptation (see [9]):

$$\frac{d}{dt}\Lambda_{c,r} = N_c\kappa_{c,r}\left[U'_{s(c)}(\Lambda_{c,r}/N_c) - \partial_r\Gamma(\Lambda)\right] + \mu_{c,r}, \quad (25)$$

where the term $\mu_{c,r}$ is non-negative and such that $\mu_{c,r}\Lambda_{c,r} \equiv 0$, and $\kappa_{c,r}$ is a positive gain parameter. We adapt the definition of the net benefit B_c per unit time for type s users streaming along routes r in some set c to the present context of uncoordinated rate control. This reads

$$B_c = \sum_{r\in c} U_s\left(\lambda_{c,r}\right) - \sum_{r\in c} \lambda_{c,r}U'_s(\lambda_{c,r}).$$

Otherwise, we assume as in the coordinated case that the evolution of the numbers N_c is characterised by the

differential equations (21). The proof of the following mirrors that of Proposition 5.1 and is omitted.

Proposition 5.2: Assume that the utility functions U_s and the penalty function Γ are continuously differentiable on their domain, that the former are strictly concave increasing, and the latter convex increasing. Assume further that $U_s'(x) \to 0$ as $x \to \infty$. Then any absolutely continuous solution $(N_c, \lambda_{c,r})$ to the system of ODE's (25-21) converges to the set of maximisers of the welfare function

$$\mathcal{W}(\Lambda, N) := \sum_{s \in \mathcal{S}} \sum_{c \subset \mathcal{R}(s)} \sum_{r \in c} N_c U_s(\Lambda_{c,r}/N_c) - \Gamma(\Lambda) \tag{26}$$

where $\Lambda_{c,r} = \lambda_{c,r} N_c$, under the constraints (3). The corresponding equilibrium rates $\Lambda_r = \sum_{c:r \in c} \Lambda_{c,r}$ are solutions of the coordinated welfare maximisation problem (6–7), with the utility function $x \to U_s(x)$ replaced by $x \to bU_s(x/b)$.

VI. Concluding remarks

We have looked at some of the properties of coordinated or uncoordinated controllers when combined with multipath routing. We have concentrated on the case with fixed-arrivals. The main findings are that without path reselection, uncoordinated control can perform poorly, and is "unfair". This resonates with the findings of [10], [13], [12], when demand is stochastic. This previous theoretical work has shown that with stochastic arrivals, uncoordinated controls can perform poorly, either giving a much smaller schedulable (stability)region than coordinated, or even when the stability regions are the same, giving poorer performance. In passing, for a simple scenario we have also given a characterisation of performance for coordinated control that does better than the greedy-least loaded routing as in Mitzenmacher for large systems. Recent work [14] shows the benefits are even more pronounced for small systems.

Early P2P systems such as Kazaar use a form of uncoordinated control without path reselection. Recent P2P filecasting applications, such a BitTorrent, implement uncoordinated congestion control using parallel TCP connections, but reselecting paths. With random path selection, whereby paths are randomly reselected, and new paths accepted if there is a net benefit, then we find first, that choosing just a small number of routes can do as well as if we tried the whole set. In addition, uncoordinated and coordinated both lead to a system optimal (welfare maximising solution), achieved in a distributed manner, *provided* the uncoordinated controllers have no RTT bias (unlike current coordinated controllers). Accepting that route re-sampling produces fair allocations *as if* all available routes were jointly used, we may consider additional dynamics of user arrivals and departures. The fluid limits for the latter system would then be the same as for the system with fair sharing, using the full set of available routes, as route resampling would take place on a fast time scale compared to the time scale of arrivals / departures in a many users limit.

This suggests good design choices for new multipath rate controllers are coordinated controllers or uncoordinated controllers with the RTT bias removed.

References

[1] N. Alon and J. Spencer, *The probabilistic method*, Wiley , 2nd Edititon, 2000.

[2] D.G. Andersen, H. Balakrishnan, F. Kaashoek and R.N. Rao, "Improving Web availiability for clients with MONET", NSDI 2005.

[3] S. Boyd and L. Vandenberghe, "Convex Optimization", CUP, 2004.

[4] B. Cohen, "Incentives built robustness in BitTorrent", in Proceeding of P2P Economics workshop, June 2003.

[5] M. Chiang and S. H. Low and A. R. Calderbank and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures", *Proc. IEEE*, December 2006.

[6] K.P. Gummadi, H. Madhyastha, S.D. Gribble, H..M. Levy, D.J. Wetherall. "Improving the reliability of Internet paths with one-hop source routing", *Proc. 6th OSDI*, 2004.

[7] H. Han, S. Shakkottai, C.V. Hollot, R. Srikant, D. Towsley. "Multi-Path TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet ", *IEEE/ACM Trans. Networking*, December 2006.

[8] F. Kelly, "Mathematical modeling of the Internet", in Mathematics Unlimited–2001 and Beyond. B. Engquist and W. Schmid, Eds. Berlin, Germany: Springer-Verlag, 2001, pp.685–702.

[9] F. Kelly, A. Maulloo, D. Tan. "Rate control in communication networks: shadow prices, proportional fairness and stability. *J. of the Operational Research Society*, **49** (1998) 237-252.

[10] P. Key and L. Massoulié. Fluid models of integrated traffic and multipath routing. *Queueing Systems*, 53(1):85–98, June 2006.

[11] F. P. Kelly and T. Voice, Stability of end-to-end algorithms for joint routing and rate control, Computer Communication Review **35:2** 5–12, 2005.

[12] L. Massoulié and P. Key. Schedulable regions and equilibrium cost for multipath flow control: the benefits of coordination. In *CISS 2006, 40th Conference on Information Sciences and Systems*, Princeton, March 2006. IEEE Information Society, IEEE.

[13] P. Key and L. Massoulié and D. Towsley. Combining multipath routing and congestion control for robustness. In *CISS 2006*, Princeton, 2006.

[14] P. Key and L. Massoulié and D. Towsley. Multipath Routing, Congestion Control and Load Balancing. in *ICASSP 2007*, Hawaii, April 2007, IEEE.

[15] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

[16] M. Kodialam, T.V. Lakshman, S. Sengupta. "Efficient and robust routing of highly variable traffic," HotNets 2004.

[17] C. de Lanois, B. Quoitin and O. Bonaventure, "Leveraging Internet path diversity and Network performances with Ipv6 Multihoming", Universite catholique de Louvain Tech. report RR 2004-06.

[18] W.-H. Wang, M. Palaniswami, S.H. Low. "Optimal flow control and routing in multi-path networks", *Performance Evaluation*, **52** (2003) 119-132.

[19] M. Mitzenmacher, A. Richa and R. Sitaram, "The Power of Two Random Choices: A Survey of Techniques and Results", in Handbook of Randomized Computing: volume 1, P. Pardalos, S. Rajasekaran and J. Rolim Eds, 2001.

[20] R. Srikant, *The Mathematics of Internet Congestion Control*, Birkhauser, 2003.

[21] Skype web site, http://skype.com/products/explained.html

[22] R. Zhang-Shen, N. McKeown. "Designing a predictable Internet backbone with Valiant load-balancing", IWQoS 2005.