MULTIPATH ROUTING, CONGESTION CONTROL AND DYNAMIC LOAD BALANCING

Peter Key
Microsoft Research
Cambridge, UK

Laurent Massoulié
Thomson Corporate Research
Paris, France

Don Towsley
University of Massachusetts
Amherst, USA

ABSTRACT

Combining transport-layer congestion control with multi-path routing is a cross-layer approach that provides performance benefits over treating the layers separately. We phrase this as an optimisation problem, examine the case of data transfers, and show how a coordinated controller gives strictly better performance than an uncoordinated controller, which sets up parallel paths. For fixed demands, and the case of random-path selection, we show how coordinated control also achieves better load balancing than greedy least-loaded path selection. We then comment on adaptive path selection.

Index Terms— Optimisation, multipath routing, congestion control, load balancing.

1. INTRODUCTION

Multipath routing, where data can potentially be sent over a number of paths offers performance and reliability benefits. Combining multipath routing with transport layer congestion control is an example of cross-layer optimisation [1, 2], and offers performance advantages over treating the layers separately. Routing determines which paths should be used, whereas rate control determines how much should be sent over each path.

Our motivating example is data transfers in a fast packet network, such as data transfers using TCP in the current Internet, and where the transfers are long enough to allow benefits for multipath routing. Following the work pioneered by Kelly et al [3], we characterise a rate-control algorithm as the solution to a utility maximisation problem, where users (or end-systems) selfishly choose paths and rates in such a way as to maximise their net utility, assuming demand is fixed. A particular rate control algorithm, is mapped to a particular utility function [4]. For example, if T is the round trip time and the user sends at rate λ, then the utility function

$$U(\lambda) = -\frac{1}{\lambda T^2} \tag{1}$$

approximately models TCP Reno's rate control. The network cost, which may reflect loss or delay is captured by a penalty function, which is a function of the load on the network.

Routing and control may be either sender driven or receiver driven. There are already a number of peer-to-peer receiver-driven multipath applications, such as Skype which maintains a number of active paths and chooses the best path, and BitTorrent [5], which uses a fixed number of paths and has a mechanism for randomly choosing an additional path and retaining the best paths.

A coordinated control actively balances load across a given set of paths, and is modelled by a single utility function per user. In contrast, an uncoordinated controller uses all available paths in parallel (for example parallel TCP connections), and achieves a much more limited form of load balancing. Previous work [6] has shown that for dynamic arrivals (stochastic demand), in general a coordinated controller has a larger schedulable region, and better performance than an uncoordinated controller.

In this paper, we concentrate on a fixed-demand scenario. We first examine the case where a fixed integer number of paths b is chosen randomly from a set of size N. We look at the worst case allocation, a measure of fairness, and show that coordinated control gives better performance, both for large N, where we quote scaling results, and for small N, where we give numerical examples. We also show this does better than greedy-least-loaded resource selection, as in Mitzenmacher [7].

We then allow users to change their routes, by resampling to choose better routes.

2. OPTIMISATION FRAMEWORK

2.1. Model and Notation

Consider a network where paths are indexed by $r \in \mathcal{R}$, and a set of user classes, indexed by $s \in \mathcal{S}$. 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 [6] 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", making use of the notation $p_r = \partial_r \Gamma(\Lambda)$.

2.2. Uncoordinated Congestion Control

We assume that class s-users try to maximise their throughputs, and that the rate control along a given route is a congestion control mechanism, such as TCP, that implicitly performs some utility maximisation, where the utility to a single user sending at rate λ_r through route r is $U_r(\lambda_r)$. For tractability, we assume that U_r is a strictly concave increasing function that is continuously differentiable on $(0, \infty)$.

Denote by N_r, $r \in \mathcal{R}(s)$, the total number of connections made by class s users along route r and N_s the total number of connections user s makes , then

$$\sum_{r \in \mathcal{R}(s)} N_r = N_s, \quad s \in \mathcal{S}, \tag{2}$$

where typically either N_r or N_s is fixed. For example, if there are N'_s class s-users (with N'_s fixed), and each class s user is restricted to using the same fixed number b of connections that can be along routes $r \in \mathcal{R}(s)$ (or if b is the cardinality of $\mathcal{R}(s)$ for all s), then $N_s := bN'_s$. 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

$$\text{Maximise} \sum_{s\in\mathcal{S}} \sum_{r\in\mathcal{R}(s)} N_r U_r(\Lambda_r/N_r) - \Gamma(\Lambda). \tag{3}$$

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

The function being optimised in (3) is a strictly concave function, over a convex feasible region; hence the problem is Strong Lagrangean and the unique maximum is attained.

2.3. Coordinated Congestion Control

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 $(0,\infty)$. N_s is the number of class s users. We can then assume that the allocation to a class s-user is $\Lambda_s = \sum_{r\in\mathcal{R}(s)} \Lambda_r$, where optimal rates Λ_r solve the following welfare maximisation

$$\text{Maximize} \sum_s N_s U_s \left(\frac{\sum_{r\in\mathcal{R}(s)} \Lambda_r}{N_s} \right) - \Gamma(\Lambda) \tag{4}$$

$$\text{over } \Lambda_r \geq 0, \quad r \in \mathcal{R}. \tag{5}$$

Note that if the utility functions U_r in (3) are path independent and coincide with U_s for all $r \in \mathcal{R}(s)$, then Jensen's inequality shows welfare for the coordinated solution (solving (4)) is at least as large the the welfare for the uncoordinated (solution to (3)). This coordinated problem is strong Lagrangean, and its solution characterized by the Kuhn-Tucker conditions

$$U_s' \left(\frac{\sum_{r\in\mathcal{R}(s)} \Lambda_s}{N_s} \right) \leq \partial_r \Gamma(\Lambda), \tag{6}$$

$$U_s' \left(\frac{\sum_{r\in\mathcal{R}(s)} \Lambda_s}{N_s} \right) < \partial_r \Gamma(\Lambda) \Rightarrow \Lambda_r = 0, \tag{7}$$

and, as a consequence, the allocation for user s only puts a non-zero allocation on paths whose price p_r is equal to the minimum price across possible routes. There may be only one or several such lowest-cost paths. Note that distributed rate control algorithms exist for all of the above optimisation problems, e.g. [8, 9].

3. STATIC ROUTE SELECTIONS AND LOAD BALANCING

We now concentrate on the following scenario: there are N resources (each corresponding to a single route r), each with unit capacity and associated penalty function

$$\Gamma_r(\Lambda_r) = \begin{cases} 0 & \text{if } \Lambda_r \leq 1 \\ \infty & \text{otherwise.} \end{cases} \tag{8}$$

There are aN users, and each user selects b resources at random from the N available, where b is an integer larger than 1; hence $R(s)$ is this set of b resources (routes). (The theoretical results assume sampling with replacement, whereas the numerical results assume sampling without replacement, which is more natural in practice). Denote by λ_{ij} the rate that user i obtains from resource j, and let A_{ij} equal 1 if user i can access resource j, and 0 otherwise. We consider the worst case rate allocation of users under two distinct bandwidth sharing scenarios, firstly when there is no coordination between the distinct b connections of each user, and secondly when each user implements coordinated multipath congestion control. In this section, we assume that $U_r = U$ for all r.

3.1. Scaling results

In the uncoordinated case, it is straightforward to show

Lemma 3.1 *The optimal allocation is independent of the choice of utility functions U_r, and is given by $\lambda_i = \sum_j \lambda_{ij}$, where $\lambda_{ij} = 1/\sum_i A_{ij}$.*

That is, if a resource serves n users, each gets rate $1/n$ from that resource. It is then possible to show [10],

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

$$\lim_{N\to\infty} \mathbf{P}\left(\min_{i=1,\ldots,aN} \lambda_i \leq (b^2+\epsilon) \frac{\log(\log(N))}{\log(N)} \right) = 1. \tag{9}$$

In other words, the worst case allocation in this scenario decreases like $\log(\log(N))/\log(N)$. This is comparable with the worst case allocation that one gets if just a single path is used: using a classical balls and bins models, eg [7], where we imagine users throwing a ball into the bin (resource), the inverse of the maximum number of balls in bin scales as $\log(\log(N))/\log(N)$ as N increases.

For coordinated congestion control, the rates λ_{ij} solve

$$\begin{aligned} \text{Maximise} \quad & \sum_{i=1}^{aN} U\left(\sum_{j=1}^{N} A_{ij}\lambda_{ij} \right) \\ \text{subject to} \quad & \sum_{k=1}^{aN} A_{kj}\lambda_{kj} \leq 1, \quad 1 \leq j \leq N \\ \text{over} \quad & \lambda_{ij} \geq 0 \end{aligned} \tag{10}$$

It is then possible to prove the following max-min fair characterisation of the optimal rates

Lemma 3.2 *Let (λ_i^*) be the optimal user rates solving the above optimisation (10), then (λ_i^*) is insensitive to the particular strictly concave, increasing utility function U chosen. Denote by $x_1 < x_2 < \cdots < x_m$ the distinct values of the λ_i^*, ranked in increasing order. Let I_1 denote the set of indices i such that $\lambda_i^* = x_1$. 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. x_1 can be found by solving the LP*

$$\textit{Maximise} \quad \lambda^*$$

$$\textit{subject to} \quad \sum_{j=1}^{N} A_k j \geq \lambda^*, \quad \sum_{k=1}^{aN} A_{kj}\lambda_{kj} \leq 1, \quad 1 \leq j \leq N$$

We can then prove the following [10]

Theorem 3.2 *If $(\lambda_i^*(N))$ is the optimal allocation for coordinated congestion control for given N, a, b, then there exists $x > 0$, that depends only on a and b, such that:*

$$\lim_{N\to\infty} \mathbf{P}\left(\min_i \lambda_i^*(N) \geq x \right) = 1. \tag{11}$$

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

$$\forall u \in (0,a], ah(u/a) + h(ux) + bu\log(ux) < 0, \tag{12}$$

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

This says that the worst-case allocation is bounded away from zero, as $N \uparrow \infty$, and strictly better than the uncoordinated allocation. This particular allocation problem has links with the load balancing work, quoted by Mitzenmacher [7],where if users arrive in a random order, and choose the lowest-loaded resource from among their b candidate ones, a 'greedy least-loaded' strategy, then with high probability the maximum resource is at most $\log\log N/\log b + O(1)$. Hence the worst-case rate scales as $1/\log(\log(N))$, whereas we do better than this. We achieve better results by actively balancing load across several available resources.

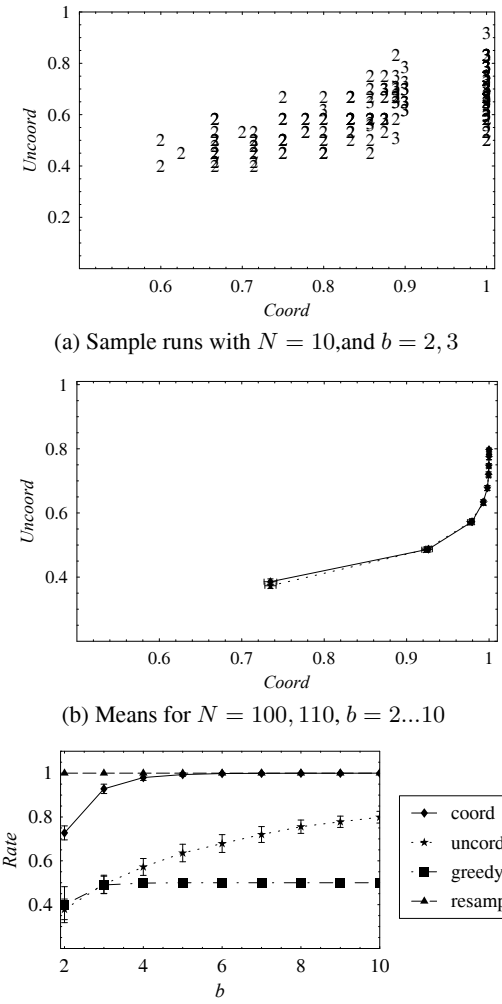
Informally, the coordinated congestion control is able to "shuffle" the load amongst the union of the sets of b resources, and for moderate b do almost as well as if each user saw the global list of N resources rather than just a subset of size b. We can think of the coordinated approach as performing a secondary optimisation once all the random choices of resources have been made. In the context of data transfers, the implemented rate control will take a finite time to adjust the rates across paths to perform the optimisation, for example a few round trip times. However, the result is applicable when the amount of data that users have to transfer is large compared to the resource capacity available, implying the transfer time is orders of magnitude larger than a round trip time.

A few remarks are in order. 1. The *resource* utilisation is the same for both coordinated and uncoordinated, that is the number of resources utilised is the number of non-empty columns of the matrix A, bounded above by N, and each resource used will serve at rate 1. 2. The coordinated allocation maximises the minimum rate any user receives (Lemma 3.2), and also minimises the expected time to transfer a unit of data (where the expectation is across users). This last fact follows from the fact that we are free to use the utility function $U(x) = -1/x$ in the coordinated optimisation, hence maximising the aggregate utilities is equivalent to minimising the aggregate download time for a fixed data unit, and hence to minimising the average.

3.2. Small System results

The above scaling results describe large system behaviour. We now look at what happens for small systems (small N) by randomly sampling sets of size b and calculating the optimal allocations. In what follows, we set $a = 1$, and vary b. The scaling results suggest that choosing b relatively small, $b = O(\log(N))$ is sufficient for coordinated control to perform with high probability a "perfect" allocation where every user receives rate one. If we pick a particular resource, then with probability $(1 - b/N)^N$ the resource is not chosen by any user, and hence "isolated". Hence the expected number of isolated resources is $N(1-b/N)^N$, and the best rate-allocation (in max-min terms) is $1-(1-b/N)^N$. If $b = \log(N)$, then $(1-b/N)^N \leq 1/N$ with the limit approached as $N \uparrow \infty$, therefore if we put $b = \log(N)$ we expect the best (max-min) rate to be $1 - 1/N$. For example, for $N = 100$, $\log 100 \approx 5$ giving an expected rate of 0.99, while for $N = 1000$, $\log 1000 \approx 7$ giving an expected minimum rate 0.999, whereas $\log(\log(N))/\log(N)$ is 0.33 and 0.28 respectively which is consistent with the results or simulation (Table 3.2 below).

Figure 1(a) shows the results of 100 sample runs for $N = 10$ and $b = 2$ and $b = 3$, where we plot the minimum rate for each sample, labelled by b. Coordinated performs much better than uncoordinated, and for $b = 3$ coordinated has in many cases achieved a 'perfect' allocation. Figure 1(b) gives means and 95% confidence intervals for $N = 100, 110$ and $b = 2...10$. There is very little difference between $N = 100$ and $N = 110$. Figure 1(c) gives means and 95% confidence intervals for $N = 100$, for coordinated, uncoordinated, greedy least-loaded strategy and coordinated with resampling, where users resample in time and move to better routes (see the next section). This last approach gives equal allocation to all users and achieves a perfect allocation provided the resampling is rich enough. We see that coordinated outperforms uncoordinated, and only a small value of b is needed for almost perfect load balancing. Interestingly the greedy load balancing does not do as well as uncoordinated - essentially in the simulations there was always some resource which has more than one user associated with it.



(a) Sample runs with $N = 10$,and $b = 2, 3$

(b) Means for $N = 100, 110$, $b = 2...10$

(c) Means for $N = 100$, $b = 2...10$

Fig. 1. Coordinated and uncoordinated control, with 100 runs.

Results for $N = 1000$ are given in Table 3.2, where $n = 10$ runs were done. Means are shown (μ) along with the standard deviation for the mean ($\sigma/\sqrt{n}$).

For $N = 100, 1000$ the rates are consistent with the scaling arguments given above. In the case of the greedy least-loaded algorithm described above, for $N = 1000$, we would expect *with high probability* a minimum load to approximate $\log b/\log\log N$ whereas we actually obtained rates of $1/3, 1/2, 1/2$ for $b = 2, 5, 10$ - for example with $b = 10$ this means there was at least one resource which had two associated users.

N	b	*coord*		*uncoord*	
		μ	$\sigma/\sqrt{n}$		$\sigma/\sqrt{n}$
1000	2	0.73	0.003	0.31	0.008
1000	5	0.993	0.001	0.55	0.004
1000	10	1.00	0.0000	0.74	0.001

Table 1. Means and Standard Deviations $N = 1000$ and varying b

4. DYNAMIC ROUTE SELECTION

In practice, there may be a large number of potential routes available to a user, and when we move outside the symmetric single capacity example of Section 3 we would like to achieve the best possible outcome without having to consider all potential routes simultaneously. It turns out we can do this by allowing users to adapt their route sets to move towards better ones. We adapt our framework, and 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, giving the constraint

$$\sum_{c \in \mathcal{C}(s)} N_c = N_s, \quad \text{for } s \in \mathcal{S}. \tag{13}$$

λ_c is the rate each of the N_c users receive, and $\Lambda_r = \sum_c N_c \lambda_c$. For coordinated control, we assume rate adaptation of the form (see [4]):

$$\frac{d}{dt}\lambda_{c,r} = \kappa_{c,r}\left[U'_{s(c)}(\lambda_c) - \partial_r \Gamma(\Lambda)\right] + \mu_{c,r}, \tag{14}$$

where $k_{c,r}$ are positive gain parameters and $\mu_{c,r} \geq 0$ satisfies $\mu_{c,r}\lambda_{c,r} \equiv 0$. Denote the net benefit per unit time for type s users streaming along routes r in some set c as B_c, then in equilibrium this is

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

since at equilibrium the price $p_r = U'_s(\lambda_c)$.

Now at instants of a Poisson process having rate $\gamma_{cc'}$ users currently using the set c will be offered a set c', and will use the set c' instead, provided they increase their *net benefit*. Given our assumptions on U, $B(\cdot)$ is an increasing function, hence equivalently, users can switch to path c' if they receive a higher *rate*. If there is a large population of users that switch according to this rule, then we can consider the deterministic limit equations

$$\frac{d}{dt}N_c = \sum_{c'} N_{c'}\gamma_{c'c}\phi(B_c - B_{c'}) - \sum_{c'} N_c \gamma_{cc'}\phi(B_{c'} - B_c) \tag{15}$$

where ϕ is the indicator function or a smooth approximation to it, and where there is an implicit separation of time scales. We can then show [10]

Theorem 4.1 *If the penalty function Γ is continuously differentiable and convex decreasing, U_s strictly concave increasing, and 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 $\gamma_{cc'}$ is positive Then any solution $(N_c, \lambda_{c,r})$ to the system of ODE's (14-15) converges to the set of maximisers of the welfare function*

$$\mathcal{W}(\lambda, N) := \sum_{s \in \mathcal{S}} \sum_{c \subset \mathcal{R}(s)} N_c U_s(\lambda_c) - \Gamma(\Lambda) \tag{16}$$

under the constraints (13). The corresponding equilibrium rates (Λ_r) are solutions of the coordinated welfare maximisation problem (4–5).

This means we are able to converge to a social optimum, as if we were simultaneously exploring all routes, but by having users limited to using a small set of routes at a particular time, using a coordinated congestion controller and then moving to better routes over time. Does this also hold for uncoordinated congestion controllers? Only if the utility functions are the same across all paths, in other words *have no RTT bias*, unlike say TCP Reno. In this case we can prove a similar welfare maximisation result, provided we replace $U_s(x)$ by $bU_s(x/b)$.

5. CONCLUDING REMARKS

We have shown the benefits of combining congestion control with multipath routing, showing that in a static setting, where random path sets are chosen, coordinated controllers outperform uncoordinated ones. In a dynamic setting, by having limited set of routes, but allowing evolution to better routes over time, we can implement a distributed welfare maximisation with coordinated controllers. For uncoordinated controllers, this is only true if there is no RTT bias in the rate controllers. This has implications for the design of practical multi-path rate control algorithms.

6. REFERENCES

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

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

[3] F. P. Kelly, A. K. Maulloo, and D. K. H Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[4] F. P. Kelly, "Mathematical modelling of the Internet," in *Proceedings of the Fourth International Congress on Industrial and Applied Mathematics*, 2000.

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

[6] Peter Key and Laurent Massoulié, "Fluid models of integrated traffic and multipath routing," *Queueing Systems*, vol. 53, no. 1, pp. 85–98, June 2006.

[7] M. Mitzenmacher, A. Richa, and R. Sitaraman, *The power of two random choices: A survey of the techniques and results*, Kluwer, 2000.

[8] F.P. Kelly and T. Voice, "Stability of end-to-end algorithms for joint routing and rate control," *Computer Communication Review*, vol. 35, no. 2, pp. 5–12, 2005.

[9] H. Han, S. Shakkottai, C.V. Hollot, R. Srikant, and D. Towsley, "Overlay TCP for multi-path routing and congestion control," *IEEE/ACM Trans. Networking*, December 2006.

[10] Peter Key, Laurent Massoulié, and Don Towsley, "Path selection and multipath congestion control," in *Proc. IEEE Infocom 2007*, May 2007.