Schedulable regions and equilibrium cost for multipath flow control: the benefits of coordination

Laurent Massoulié & Peter Key
Microsoft Research, 7 J J Thompson Avenue, Cambridge CB3 0FB, UK

***Abstract*— In this paper we consider deterministic differential equation models for the varying number of flows in a network. These arise naturally as limits of stochastic models under joint scaling of flow arrival rates and network capacities. We compare these dynamics under (i) coordinated multipath routing and (ii) parallel, uncoordinated routing. We show that for identical traffic demands, parallel uncoordinated routing can be unstable while balanced multipath routing is stable. In other words, coordination can strictly increase the schedulable region, that is the set of demand vectors for which the system is stable. We also show that, even when uncoordinated multipath routing stabilises the system, coordination can bring further benefits, as it naturally minimises network costs at equilibrium.**

I. INTRODUCTION

In all transportation networks where the transported goods can be split, there is a clear case for achieving transportation along multiple paths rather than a single one, as this potentially increases the transport capacity. The Internet is no exception, and indeed, a significant part of today's Internet data downloads proceed along multiple paths: peer-to-peer file sharing downloads from popular systems such as BitTorrent or eDonkey let files be sent from several sources in parallel.

Such multipath downloads typically rely on individual TCP connections for each path. Proposals have been made recently to extend the existing TCP protocol to let it orchestrate parallel paths directly, in contrast to the use of uncoordinated, single path TCP connections [10], [4], [3].

The question we address in the present paper is which kind of coordination is desirable between such multiple routes, if any. More precisely, we consider an optimisation framework for the definition of bandwidth allocation objectives, coordinated or not. We then focus on flow-level dynamics, with flow arrivals and departures, assuming a *fluid* scaling corresponding to large capacities and arrival rates.

We compare the performance of coordinated and uncoordinated multipath allocations by evaluating for both schemes:

(i) the *schedulable region*, that is the set of demand parameters for which the dynamics are asymptotically stable, and

(ii) the *equilibrium cost*, that is the cost of network resource usage at equilibrium when the dynamics do reach equilibrium.

Our findings are as follows. For the specific model of coordination we consider, the schedulable region is maximised, while counter-examples are provided where uncoordinated controllers can achieve strictly smaller schedulable region. We then identify topologies of interest where uncoordinated controllers also achieve maximal stability region. We then show that the equilibrium cost is minimised for coordinated controllers, and provide counter-examples illustrating this does not hold for uncoordinated controllers.

II. MODELING ASSUMPTIONS

We assume flows can be of several types indexed by $s \in \mathcal{S}$, where $\mathcal{S}$ is a finite set, and denote by

n_s the number of type s flows at a given time. We assume a set of paths, or routes, $\mathcal{R}$ is given. Flows of type s can make use of paths $r \in \mathcal{R}_s$ for some corresponding path set $\mathcal{R}_s \subset \mathcal{R}$. Without loss of generality, we may and shall assume that these sets are disjoint. We shall denote by $s(r)$ the type of flows using route r.

Finally, we assume there is a cost function $\Gamma : \mathbb{R}_+^{\mathcal{R}} \to \mathbb{R}_+ \cup \{+\infty\}$ such that, when the flow along path r is Λ_r, $r \in \mathcal{R}$, the total network cost is given by $\Gamma(\Lambda)$.

Throughout, Γ is assumed to be non-decreasing and convex. A typical example for the cost function Γ is as follows. Here, network resources consist in links $\ell \in \mathcal{L}$, each with capacity C_ℓ, and a path r consists in a series of such links. Write $\ell \in r$ to denote that path r traverses link ℓ. Then, specifying Γ as

$$\Gamma(\Lambda) = \begin{cases} 0 & \text{if } \sum_{r:\ell \in r} \Lambda_r \le C_\ell,\ \ell \in \mathcal{L}, \\ +\infty & \text{otherwise,} \end{cases}$$

captures the link capacity constraints, putting zero cost on feasible path rate allocations, and infinite cost on infeasible allocations.

We now describe bandwidth allocation criteria.

A. Uncoordinated bandwidth allocation criterion

To each path r is associated a utility function $U_r : \mathbb{R}_+ \to \mathbb{R}_+$, assumed increasing and strictly concave. Then, given the number of flows n_s of each type $s \in \mathcal{S}$, the uncoordinated path bandwidth allocations $\{\Lambda_r\}_{r\in\mathcal{R}}$ are defined as:

$$\Lambda \in \text{argmax} \left\{ \sum_{r \in \mathcal{R}} n_{s(r)} U_r \left(\frac{\Lambda_r}{n_{s(r)}} \right) - \Gamma(\Lambda) \right\}. \tag{1}$$

Note that this corresponds to the classical utility maximisation criterion introduced in the context of bandwidth sharing in [**?**], with the specificity that the number of type r flows is in fact exactly $n_{s(r)}$, that is the number of flows of the corresponding type. Given the solution Λ to this optimisation problem, $\Lambda_r/n_{s(r)}$ would then represent the bandwidth obtained by any type s flow along its route r. For a discussion of utility functions U_r modeling standard TCP Reno behaviour, see e.g. [**?**], [7]. In particular, given a parameter $\alpha > 0$ and positive route weights w_r, the choice

$$U_r(x) = \begin{cases} w_r \frac{x^{1-\alpha}}{1-\alpha} & \text{if } \alpha \neq 1, \\ w_r \log(x) & \text{if } \alpha = 1 \end{cases} \tag{2}$$

gives rise to the so-called weighted alpha-fair sharing, introduced in [8]. TCP's behaviour is well approximated by taking $\alpha = 2$, and $w_r = 1/T_r^2$ where T_r is the round-trip time of route r.

Thus, each type s-flow would receive under this model a bandwidth allocation equal to $\sum_{r\in\mathcal{R}_s} \Lambda_r/n_s$. Note that strict concavity of the objective function being maximised in (1) ensures that the solution is unique if it exists, except when some coordinates n_s are zero. In the latter case, we choose the specific solution which has $\Lambda_r = 0$ for all routes corresponding to sources s with $n_s = 0$. In all examples we shall consider, existence will hold trivially. For example, if the cost function Γ derives from link capacities as above, then the optimisation is effectively restricted to a bounded set, which ensures existence.

B. Coordinated bandwidth allocation criterion

Here we assume that there are utility functions U_s attached to each traffic source rather than to each route. Again the U_s are increasing and strictly concave. The path bandwidth allocations Λ_r are then defined as solutions to the optimisation problem

$$\Lambda \in \text{argmax} \left\{ \sum_{s\in\mathcal{S}} n_s U_s \left(\frac{\sum_{r\in\mathcal{R}_s} \Lambda_r}{n_s} \right) - \Gamma(\Lambda) \right\}. \tag{3}$$

As before, given a solution Λ to this optimisation problem, the resulting allocation to each type s flow is specified as $\sum_{r\in\mathcal{R}_s} \Lambda_r/n_s$. The above optimisation problem is not necessarily strictly concave and there may exist multiple solutions. However, the resulting flow allocation is uniquely defined for all s such that $n_s = 0$. Indeed, upon defining the modified cost function $G : \mathbb{R}_+^{\mathcal{S}} \to \mathcal{R}_+ \cup \{+\infty\}$ at $x \in \mathbb{R}_+^{\mathcal{S}}$ as the supremum of $\Gamma(\Lambda)$ over $\Lambda \in \mathbb{R}_+^{\mathcal{R}}$ such that

$$\sum_{r\in\mathcal{R}_s} \Lambda_r = x_s, s \in \mathcal{S},$$

it holds that for any solution Λ to (3), the corresponding total flow type rates $x_s = \sum_{r\in\mathcal{R}_s} \Lambda_r$ satisfy

$$x \in \text{argmax}\left\{\sum_{s\in\mathcal{S}} n_s U_s(x_s/n_s) - G(x)\right\}. \quad (4)$$

Strict concavity of the above optimisation problem in the variables x_s then ensures uniqueness of the solution.

This type of coordination has been considered for instance in [3] and [4], where congestion controllers aiming to achieve such bandwidth sharing in a distributed manner are proposed.

C. Dynamics

Let positive flow arrival rates ν_s and average flow sizes μ_s^{-1} be given for each flow type $s \in \mathcal{S}$. The parameter ν_s could be thought of the rate of a Poisson process counting arrivals of new type s flows, and μ_s as the parameter of an exponential distribution from which type s flow sizes are sampled.

We shall not deal directly with this stochastic model, but rather consider deterministic, *fluid* dynamics, that can be considered as the result of a simultaneous scaling of system capacities and arrival rates in the original stochastic model. A more detailed discussion of this type of rescaling is provided in [5]. A detailed proof of the convergence of stochastic process trajectories to the fluid model trajectories is provided in [6] in the case of a single resource.

The fluid dynamics we shall subsequently deal with are as follows:

$$\frac{d}{dt} n_s = \nu_s - \mu_s \sum_{r\in\mathcal{R}_s} \Lambda_r(n),\ s \in \mathcal{S}. \quad (5)$$

In the above, $\Lambda(n)$ will denote a solution of either (1) or (3) according to whether we consider the uncoordinated or coordinated case, where we make explicit the dependency on the vector of flow numbers, $n = (n_s)_{s\in\mathcal{S}}$.

III. Stability properties

We now turn to the asymptotic stability properties of the above dynamics.

A. Asymptotic stability with coordination

Define the load ρ_s of type s traffic as ν_s/μ_s. Define the *domain* of the cost function G, denoted $\text{dom}(G)$, as the set of vectors $x \in \mathbb{R}_+^{\mathcal{S}}$ such that $G(x) < \infty$. Assume that for all $\epsilon > 0$, all $s \in \mathcal{S}$, and all subgradient $G'(\epsilon e_s)$ of G at ϵe_s, where e_s is the s-th unit vector in $\mathbb{R}_+^{\mathcal{S}}$, one has

$$\lim_{x\to\infty} U_s'(x) < G_s'(\epsilon e_s).$$

This assumption guarantees that the dynamics (5) are pushed away from the boundary of the orthant $\mathbb{R}_+^{\mathcal{S}}$. We then have the following:

Theorem 1. *Assume that the load vector ρ belongs to the interior of* $\text{dom}(G)$*, and that there exists $\delta > 0$ such that*

$$U_s'(\delta) > G_s'(\rho) \quad (6)$$

for any vector $G'(\rho)$ that is a sub-gradient of G at ρ. Let

$$f_s(x) = \int_0^x U_s'\left(\frac{\rho_s}{y}\right) dy. \quad (7)$$

Then for any sub-gradient $G'(\rho)$ of G at ρ, the function

$$L(n) = \sum_{s\in\mathcal{S}} \frac{1}{\mu_s} \left\{f_s(n_s) - n_s G_s'(\rho)\right\} \quad (8)$$

is a Lyapunov function for the dynamics (5), and the solution to (5) converges to the set of equilibrium points characterised by

$$n_s^* = \frac{\rho_s}{U_s'^{-1}(G_s'(\rho))},\ s \in \mathcal{S}\cdot \quad (9)$$

where $G'(\rho)$ spans the set of sub-gradients of G at ρ.

This result is established in [5], using a proof technique of [2]. The fact that, for this model of coordination, the argument of [2], originally designed for single-path congestion control, also applies to such coordinated multipath congestion control had previously been noted in [3].

We may use this result to characterise the set of load vectors ρ for which stability holds, that is the *schedulable region* of the network. The theorem says that this schedulable region will include any vector ρ such that $\rho \in \text{int}(\text{dom}(G))$, and $G_s'(\rho) <$

$U_s'(0^+)$. Consider for the sake of illustration the case where the utility functions U_s as in (2), that is $U_s'(x) = w_s x^{-\alpha}$ for some positive w_s and α. In that case, $U_s'(0+) = +\infty$, and thus the theorem guarantees that the schedulable region contains the interior of the domain of G.

It is interesting to re-cast this result in terms of the original cost function, specified in terms of path rates:

Corollary 1. *Assume that for all $r \in \mathcal{R}$, all $\epsilon > 0$, and all subgradient $\Gamma'(\epsilon e_r)$ of Γ at ϵe_r, it holds that*

$$\lim_{x\to\infty} U_{s(r)}'(x) < \Gamma_r'(\epsilon e_r).$$

Let ρ^* denote the path load vector $(\rho_r^*)_{r\in\mathcal{R}}$ that minimises $\Gamma(\rho)$ over all path load vectors $(\rho_r)_{r\in\mathcal{R}}$ such that

$$\rho_s = \sum_{r\in\mathcal{R}_s} \rho_r, \; s \in \mathcal{S}.$$

Assume further that the vector ρ^* is in the interior of dom(Γ), and that there exists $\delta > 0$ such that for all s and all $r \in \mathcal{R}_s$ such that $\rho_r^* > 0$, $U_s'(\delta) > \Gamma_r'(\rho^*)$, for all subgradients $\Gamma'(\rho^*)$. Then the dynamics (5) eventually reaches a limiting set $\mathcal{N}$. Furthermore, for any n^* in the limiting set $\mathcal{N}$, $\{\Lambda(n^*)\}_{r\in\mathcal{R}}$ achieves a cost-optimal split of the offered loads ρ_s, that is:

$$\Gamma(\Lambda(n^*)) = \min\{\Gamma(\Lambda), \; \sum_{r\in\mathcal{R}_s} \Lambda_r = \rho_r, \; s \in \mathcal{S}\}.$$

For details, see [5].

B. Instability in the uncoordinated case

We now give a counter-example which illustrates that, in general, the use of parallel, uncoordinated connections reduces the stability region.

Consider the triangle network of Figure 1. Flows between any two pairs of nodes (say B-C) can go along the one-link route (B-C) between the nodes, or use the alternate two-hop route (B-A-C). All three links are assumed to be of unit capacity. We denote by ρ_A the load to be carried from B to C, and call the corresponding file transfers type-A, and symmetrically ρ_B and ρ_C for file transfers of types B and C. Standard manipulations show that, when both direct and indirect routes are allowed, the capacity region is described by

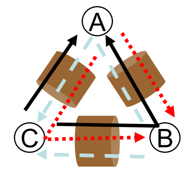

Fig. 1. Example network where parallel uncoordinated connections lead to inefficiency

$$\begin{cases} \rho_B + \rho_C \le 2, \\ \rho_C + \rho_A \le 2, \\ \rho_A + \rho_B \le 2. \end{cases} \tag{10}$$

In the symmetric case when all three offered loads coincide, the stability condition reads $\rho \le 1$. In this case, stability can be achieved without using the alternate routes.

Let us now see what happens when each file transfer uses two connections, one direct and one indirect. For definiteness assume that the allocation to each connection is α-fair, with equal weights for all connections. Let n_A (respectively n_B, n_C) denote the number of type-A (respectively, type-B, type-C) file transfers. Then file transfers of type i proceed at rate x_i, $i = A, B, C$, where

$$\begin{cases} x_A &= \left(\frac{1}{p_A}\right)^{1/\alpha} + \left(\frac{1}{p_B+p_C}\right)^{1/\alpha}, \\ x_B &= \left(\frac{1}{p_B}\right)^{1/\alpha} + \left(\frac{1}{p_A+p_C}\right)^{1/\alpha}, \\ x_C &= \left(\frac{1}{p_C}\right)^{1/\alpha} + \left(\frac{1}{p_A+p_B}\right)^{1/\alpha}, \end{cases}$$

and the p_i's are the Lagrange multipliers associated with the link capacity constraints. They are uniquely determined by

$$\begin{cases} n_A\left(\frac{1}{p_A}\right)^{1/\alpha} + n_B\left(\frac{1}{p_A+p_C}\right)^{1/\alpha} + n_C\left(\frac{1}{p_A+p_B}\right)^{1/\alpha} = 1, \\ n_B\left(\frac{1}{p_B}\right)^{1/\alpha} + n_C\left(\frac{1}{p_A+p_B}\right)^{1/\alpha} + n_A\left(\frac{1}{p_B+p_C}\right)^{1/\alpha} = 1, \\ n_C\left(\frac{1}{p_C}\right)^{1/\alpha} + n_A\left(\frac{1}{p_B+p_C}\right)^{1/\alpha} + n_B\left(\frac{1}{p_A+p_C}\right)^{1/\alpha} = 1. \end{cases} \tag{11}$$

The fluid equations, in the case of symmetric loads, then read

$$\frac{d}{dt}n_i(t) = \nu - \mu n_i x_i(\mathbf{n}(t)), \quad i = A, B, C. \quad (12)$$

We have the following:

Proposition 1. *Define ρ as λ/μ, and*

$$\rho^* := \frac{1 + 2^{-1/\alpha}}{1 + 2^{1-1/\alpha}}. \quad (13)$$

The solution $\mathbf{n}(t)$ to the system of differential equations (12) diverges to infinity whenever $\rho > \rho^*$. In particular, when $\alpha = 2$, the system is unstable provided $\rho > 1/\sqrt{2} \approx 0.71$.

Conversely, the solution $\mathbf{n}(t)$ decreases to zero in time at most $\theta[n_A(0) + n_B(0) + n_C(0)]$ for a suitable constant $\theta > 0$ whenever $\rho < \rho^*$.

For a proof, the reader can consult [5].

In contrast, for symmetric loads on the same network, by the previous results, the fluid trajectories converge to a finite equilibrium provided $\rho < 1$, hence a discrepancy between the schedulable regions with and without coordination.

C. Bipartite networks

Consider the case where each file transfer type s can use several paths $r \in \mathcal{R}_s$, and each such path consists of a single link, as illustrated by Figure 2. The resources are then a collection of links, denoted by $\ell \in \mathcal{L}$, and $\Gamma_\ell(\cdot)$ is the cost function associated with link ℓ. For definiteness, let us first consider sharp capacity constraints: $\Gamma_\ell(x) = 0$ if $x \le C_\ell$, and $+\infty$ otherwise. As usual, denote by ρ_r the offered load due to type-r users.

The stability condition, in this context, has the simple form:

$$\sum_{r \in \mathcal{S}} \rho_r < \sum_{\ell \in \mathcal{L}(\mathcal{S})} C_\ell, \quad \mathcal{S} \subset \mathcal{R}, \quad (14)$$

where the subset of links $\mathcal{L}(\mathcal{S})$ is defined to be the union of the sets $\mathcal{P}(r)$ for all $r \in \mathcal{S}$. The conditions with non-strict inequalities instead of strict ones are clearly necessary for the existence of a feasible allocation of each load ρ_r to the links in $\mathcal{P}(r)$. The fact that these conditions are also sufficient is known as Hall's theorem (see e.g. [1], p.77) in the case where

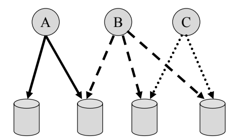

Fig. 2. Example network with 1-hop routes: parallel unccordinated connections achieve maximal stability region

the ρ_r and the C_ℓ all equal 1. This extends (i) to integral loads and capacities by splitting each traffic source and link into components with unit capacity, (ii) to rational loads and capacities by rescaling, and (iii) to arbitrary positive loads and capacities by continuity.

Denoting as before by n_r the number of type r-users, and by $U_{r\ell}$ the utility function used to determine their allocated rate via link ℓ, the allocation $x_r(\mathbf{n})$ to type-r users is then specified as

$$x_r = \sum_{\ell \in \mathcal{P}(r)} x_{r\ell}, \quad (15)$$

where the $x_{r\ell}$ maximise

$$\sum_r n_r \sum_{\ell \in \mathcal{P}(r)} U_{r\ell}(x_{r\ell}) - \sum_{\ell \in \mathcal{L}} \Gamma_\ell \left(\sum_{r : \ell \in \mathcal{P}(r)} n_r x_{r\ell} \right). \quad (16)$$

We then have the following result, proven in [5]:

Proposition 2. *Assume that for all links ℓ, and all user types r, s such that $\ell \in \mathcal{P}(r) \cap \mathcal{P}(s)$, the ratio $\frac{U'_{r\ell}(x)}{U'_{s\ell}(x)}$ is bounded away from zero and infinity, uniformly in $x \in \mathbb{R}_+$. Then, under the stability condition (14), the solutions to the system of differential equations*

$$\frac{d}{dt}n_r(t) = \nu_r - \mu_r n_r x_r \quad (17)$$

where x_r is specified by (15–16) return to zero in time at most $\theta \sum_{r \in \mathcal{R}} n_r(0)$ for some suitable constant θ.

Even for network topologies as in the previous proposition, where parallel connections achieve stability whenever synchronised parallel connections do, one may still prefer the latter allocation to the former. For instance, consider a single type of users, who can simultaneously access two resources, ℓ_1 and ℓ_2, with associated cost functions Γ_1 and Γ_2 respectively. We know from Corollary 1 that, using coordinated multiple connections, the load ρ is at equilibrium split into ρ_1 and ρ_2 so that the cost $\Gamma_1(\rho_1) + \Gamma_2(\rho_2)$ is minimised.

In contrast, in the case of parallel, uncoordinated connections, based on respective utility functions U_1, U_2, at equilibrium the loads ρ_1 and ρ_2 are now specified by the fixed point equations in the variables:

$$\begin{cases} \rho_1 + \rho_2 = \rho, \\ \rho_i = nx_i, \ i = 1, 2, \\ U_i'(x_i) = \Gamma_i'(nx_i) + \beta_i, \ i = 1, 2, \end{cases}$$

where β_i is the Lagrange multiplier associated with the constraint $x_i \geq 0$ in the optimisation problem

$$\text{Maximise} \quad n\,[U_1(x_1) + U_2(x_2)] - \Gamma_1(nx_1) - \Gamma_2(nx_2).$$

Consider for instance the case where $U_1(x) = U_2(x) = \log(x)$, and $\Gamma_i(x) = p_i x$, and assume $p_1 < p_2$. In the coordinated case, we obtain $\rho_1 = \rho$, $\rho_2 = 0$, and a corresponding cost of ρp_1.

In the uncoordinated case we obtain $\rho_1 = \rho p_2/(p_1 + p_2)$, $\rho_2 = \rho p_1/(p_1 + p_2)$ and a corresponding cost of $\rho p_1[2p_2/(p_1 + p_2)]$, larger than the optimal cost by a factor of $2p_2/(p_1 + p_2)$. This illustrates the fact that, even when stability is not lost, lack of coordination can still be detrimental.

IV. Conclusion

In this paper we have analysed the schedulable region achieved by coordinated multipath flow controllers. We have seen that, for utility functions such as these in (2), coordinated controllers achieve maximal schedulable region. We ahve given an example illustrating that uncoordinated multipath congestion control can produce a srtictly smaller schedulable region. Other realistic topologies where this happens can easily be constructed. Thus, this gives a first argument for preferring coordinated over uncoordinated multipath congestion controllers. We have then shown that on bipartite topologies, uncoordinated controllers achieve the same maximal stability region as coordinated ones, when network cost functions only reflect strict capacity constraints. We have then illustrated on an example that, when cost functions capture non-trivial network costs, even when uncoordinated controllers are stable, they achieve strictly worse equilibrium cost than coordinated controllers, while these in all circumstances achieve optimal equilibrium network cost, for given demands. This provides a second argument in favour of coordinated congestion controllers.

References

[1] B. Bollobás, *Modern Graph Theory*, Springer, 2002.

[2] T. Bonald and L. Massoulié, Impact of fairness on Internet performance. In *Proceedings of ACM SIGMETRICS 2001*.

[3] H. Han, S. Shakkottai, C.V. Hollot, R. Srikant and D. Towsley, Overlay TCP for Multi-Path Routing and Congestion Control, submitted to IEEE/ACM Transactions on Networking.

[4] 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.

[5] P. Key and L. Masssoulié, Fluid models of integrated traffic and multipath routing, to appear in Queueing Systems: Theory and Applications, special issue on queueing models for fair resource sharing.

[6] S. Kumar and L. Massoulié, Fluid and diffusion approximations of an integrated traffic model, Microsoft Research Technical Report MSR-TR-2005-160. Available at http://research.microsoft.com/users/lmassoul/MSR-TR-2005-160.ps .

[7] S. Kunnyur and R. Srikant, End-to-end congestion control schemes: utility functions, random losses and ECN marks. IEEE INFOCOM, 2000.

[8] J. Mo and J. Walrand, Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking* **8**, 556–567, 2000.

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

[10] M. Zhang, J. Lai, A. Krishnamurthy, L. Peterson and R. Wang. A Transport Layer Approach for Improving End-to-End Performance and Robustness Using Redundant Paths, USENIX 2004.