Fluid models of integrated traffic and multipath routing

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

December 1, 2005

Abstract

In this paper we consider a stochastic model describing the varying number of flows in a network. This model features flows of two types, namely file transfers (with fixed volume) and streaming traffic (with fixed duration), and extends the model of Key, Massoulié, Bain and Kelly [27] by allowing more general bandwidth allocation criteria. We analyse the dynamics of the system under a fluid scaling, and show Lyapunov stability of the fluid limits under a natural stability condition. We provide natural interpretations of the fixed points of these fluid limits.

We then compare the fluid dynamics of file transfers under (i) balanced 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.

Finally, we identify multi-dimensional Ornstein-Uhlenbeck processes as second-order approximations to the first-order fluid limit dynamics.

1 Introduction

The behaviour of large-scale communication networks, such as the current Internet, has proved to be a rich seam to mine for the research community, with research spanning many disciplines and with many questions still unanswered. In this paper we seek to describe such a network at the flow level, using stochastic analysis and optimisation theory. In particular, we study networks that are large and scaled in such a way that first-order behaviour is well-described by the deterministic limits of the underlying stochastic process, and where second-order behaviour is characterised by certain diffusion processes.

The models we describe apply generally to a variety of resource allocation situations, however they are motivated by high-speed packet networks, where there is some notion of flow between end-points in the network, and where a flow comprises a number of packet transfers. For example, think of TCP flows in the current Internet. TCP is used to carry so-called *elastic* traffic, which can adapt its rate to the underlying network conditions. Much current research has focussed on the design of rate-control algorithms and their behaviour under essentially static traffic patterns, where the load on the network is determined by a fixed, a priori given, number of flows; see for

example [40]. In contrast, we consider flow-level models under stochastic load, building on the foundations laid by Massoulié and Roberts [32], and others [13, 4, 6].

The paper falls into two parts. First, we characterise the performance of an integrated network with heterogeneous traffic of two types, which we label 'streaming' or 'file transfers'. In this part of the paper we summarise previous literature and generalise the results of [27]. A specific generalisation covers multipath routing and is sufficiently important to be treated separately, which we do in the second part of the paper. There, we explore differences between so-called "coordinated" and "uncoordinated" multipath routing; for ease of exposition we elide the streaming traffic in these comparisons. For both parts of the paper the analysis uses fluid limits and bandwidth sharing models. At the end of the paper we outline how diffusion models can be used to describe behaviour of fluctuations around the fluid limits.

By 'file transfers', we mean flows which have a given volume to transfer: the volume may be random, but is independent of the state of the network. Conversely, 'streaming traffic' has a given duration (holding time); the holding time may be random but is independent of network conditions.

In the current internet, streaming traffic may be carried by TCP or UDP. At the time of writing, TCP is the dominant transport protocol, measurements on a backbone [21] show TCP comprises 70% of the flows, and 90% when measured by volume, with UDP the main alternative protocol. Although streaming volumes are currently relatively small, we would like to explore what happens in different scenarios.

How streaming traffic (or more generally UDP) should co-exist with file-transfers is a vexed question. Some see UDP as inherently problematic. UDP has no flow control, which has led some authors to argue that streaming traffic should be "TCP-friendly" [17], while others have argued that some form of distributed or end-point admission control is necessary to assure some form of quality of service [23, 10, 3]. Analyses of traffic integration models, assuming either prioritization in favour of streaming traffic, admission control for streaming traffic, or fair sharing between all flows, can be found in references [2, 15, 7, 35].

The emergence of Voice Over IP (VOIP) provokes questions about the integration or not of different types of traffic.

The analysis of streaming traffic on its own gives rise to a product-form solution under certain reasonable assumptions, a form which is preserved under certain types of call admission control [23]. Moreover the limiting behaviour as the size of the system grows leads naturally to a non-degenerate limit for the (scaled) number of connections. In contrast, a similar scaling applied to just file transfer traffic leads to a distribution that is either unstable or has mean zero; it has been suggested [12] that such a model is flawed, lacking any self-limiting behaviour. We shall see that this criticism is avoided when the two types of traffic are mixed, and that the presence of even a small amount of streaming traffic has a stabilising effect.

Multipath routing has potential benefits in terms or performance and reliability, and recent work [20, 25] has shown that stable multi-path rate-adaptive algorithms can be constructed. One possible application scenario is in overlay networks, and another is where hosts are multi-homed. We show the benefits at the flow level of such routing, where there is some coordination so that each flow is able to optimally spread its load amongst different routes, and compare this with the case

of limited coordination: the flow is spread amongst the different routes but they act independently to transfer the data. The latter is simpler to implement with existing transport protocols, but has potential disadvantages in terms of both efficiency and stability.

The outline of the paper is as follows. In section 2 we describe the bandwidth sharing models, gradually increasing complexity. In all cases, we are able to express the sharing models in terms of an optimisation, where flows have some notional associated utility function, which can be thought of as an abstraction that various rate-allocation schemes implicitly follow. We first start with the so-called α-fair sharing scheme, for which TCP-friendliness is a special case, for a general network with fixed routing and fixed capacity constraints. We extend this in a number of ways, first by allowing more generalised utility functions, and secondly by relaxing the assumptions on sharp capacity constraints, to allow for feedback from the network that may be signalled by packet loss, packet delay, or explicitly, perhaps via packet marking. We then discuss more general routing schemes, allowing traffic to be spread across routes in a coordinated way (multi-path routing) or a relatively simple way (parallel routing).

In section 3, we describe the flow-level stochastic model, where flows arrive as a Poisson process, and depending on their type have either a fixed mean duration or a fixed mean volume. In section 4 we explore the limiting dynamics under a large systems scaling: we show that there is a unique invariant point for the scaled process, and give an interpretation in terms of a 'reduced capacity'. Informally, the file transfers place an irreducible load on the network, if we remove this and a notional associated capacity, then what is left, the reduced capacity, is shared out amongst the streaming traffic, whose allocation also determines that of the file-transfers. Although we only consider the situation where streaming flows always join and receive a fair bandwidths share, the performance predictions derived from the fluid limits also apply to a different model of integration, where streaming flows join the system under a dynamic admission control policy, and use a fixed capacity if admitted; see [27] for details.

We then show that, under natural stability conditions, the dynamics are asymptotically stable in the sense of Lyapunov. For flexible routing, we show that multi-path routing spreads the load out amongst those routes which have the same minimum end-to-end cost, thus balancing load optimally between multiple paths.

We give examples of multipath routing and parallel routing in Section 5, where we focus on the case where now there are only file-transfers. We provide examples of topologies for which multipath routing has a strictly larger stability region than parallel routing.

In section 6 we look at second-order properties of the scaled processes, and show that under natural assumptions when streaming flows are present, the deviations of the scaled processes about their mean can be described as coupled Ornstein-Uhlenbeck processes. We draw some conclusions in Section 7.

2 Bandwidth Sharing Criteria

We now review a number of network bandwidth allocation criteria for competing flows, presented in the order of increasing generality. Throughout this section, flows can be of several types, indexed by $r \in \mathcal{R}$, where $\mathcal{R}$ is a countable set, and N_r denotes the number of type r-flows, for N_r a non-negative integer.

2.1 (w, α) fairness, fixed routes and sharp capacity constraints

Consider a network with resources labelled by $j \in J$. For the moment, let a flow of type r identify a non-empty subset of J (which can be interpreted as the set of resources used by a flow on route r). Set $A_{jr} = 1$ if resource j lies on route r (i.e. $j \in r$), and $A_{jr} = 0$ otherwise. We assume positive finite capacities $(C_j, j \in J)$. Given a fixed parameter $\alpha \in (0, \infty)$ and strictly positive weights $(w_r, r \in R)$, we suppose that the bandwidth allocation to each of the N_r type r flows is x_r, where $\mathbf{x} = (x_r, r \in \mathcal{R})$ is a solution to the following optimization problem:

$$\text{maximise} \qquad \sum_{r \in \mathcal{R}} w_r N_r \frac{x_r^{1-\alpha}}{1-\alpha} \tag{1}$$

$$\text{subject to} \quad \sum_{r \in \mathcal{R}} A_{jr} N_r x_r \leq C_j, \quad j \in J \tag{2}$$

$$\text{over} \qquad x_r \geq 0, \quad r \in \mathcal{R}. \tag{3}$$

Call the resulting allocation a weighted α-fair allocation [34].

The strict concavity of the objective function (1) as a function of $(x_r, r : N_r > 0)$ and the convexity of the constraints ensures that for any solution $\mathbf{x}$ to (1–3), the component x_r is uniquely determined if N_r is positive. The solution to the problem (1–3) can be expressed in terms of Lagrange multipliers $(p_j, j \in J)$ as follows

$$x_r = \left(\frac{w_r}{\sum_j p_j A_{jr}} \right)^{1/\alpha}, \tag{4}$$

where there is one non-negative multiplier p_j for each of the capacity constraints (2), which satisfy the constraint qualification conditions

$$p_j \geq 0, \quad p_j \left(C_j - \sum_r A_{jr} N_r x_r \right) = 0, \quad j \in J. \tag{5}$$

These are the so-called complementary slackness conditions. This representation in terms of Lagrange multipliers holds because the above optimisation problem satisfies the so-called Slater conditions, i.e. there exists a vector $\mathbf{x} \in \mathbb{R}_+^{\mathcal{R}}$ such that for all $j \in J$, constraint (2) is met at $\mathbf{x}$, and is met with strict inequality for j such that the constraint is not affine ; see e.g. [8], p.226 or [38], Theorem 28.2, p.277.

When $w_r = 1, r \in \mathcal{R}$, the cases $\alpha \to 0$, $\alpha \to 1$ and $\alpha \to \infty$ correspond respectively to an allocation which achieves maximum throughput, is *proportionally fair* or is *max-min fair* [6, 34]. Weighted α-fair allocations provide a tractable theoretical abstraction of decentralized packet-based congestion control algorithms such as TCP.

If $\alpha = 2$ and w_r is the reciprocal of the square of the round trip time on route r, then the formula (4) is a version of the *inverse square root law* familiar from studies of the throughput of TCP connections [16, 33, 36]. A flow carrying streaming traffic is termed *TCP-friendly* if, *inter alia*, it adapts its rate to correspond with the steady-state rate of a TCP connection, usually characterised in terms of a version of the inverse square root law [17].

The relations (1–5), and more refined versions of these relations, can be solved to give predictions of throughput, given the numbers of flows N present [1, 11, 19, 39]. Given N, network performance along different routes can be predicted. But what determines the behaviour of N? One aim of this paper is to better understand how the behaviour of N is influenced by the mix of traffic types present, and how N is affected if we allow more flexible routing.

2.2 Generalised fairness: bandwidth utility functions

Our first, simple, extension of the above framework is to generalise the objective function in the above optimisation problem, while still identifying flow types r with subsets of resources $j \in J$. We now let the rate allocation x_r to type r flows be the solution of

$$\text{maximise} \qquad \sum_{r \in \mathcal{R}} N_r U_r(x_r) \tag{6}$$

$$\text{subject to} \quad \sum_{r \in \mathcal{R}} A_{jr} N_r x_r \le C_j, \quad j \in J \tag{7}$$

$$\text{over} \qquad x_r \ge 0, \quad r \in \mathcal{R}. \tag{8}$$

where U_r is an increasing, strictly concave function on $\mathbb{R}_+$ for all $r \in \mathcal{R}$. Here U_r is interpreted as the utility function of type r flows, $U_r(x_r)$ then representing the value of a type r flow proceeding at rate x_r. Thus the allocation $\mathbf{x}$ maximises the total utility under the network capacity constraints. Our assumptions ensure uniqueness of the allocation vector $\mathbf{x}$, and for technical convenience we also assume in addition that U_r is twice differentiable on $(0, +\infty)$.

The optimisation (1–3) of the previous section corresponds to the special case

$$U_r(x) = w_r x^{1-\alpha}/(1-\alpha).$$

The extension considered here is interesting because there are utilities of interest which may not be α-fair, for example a more refined analysis of TCP has suggested that it might allocate bandwidth according to the above criterion, with $U_r(x) = \tau_r^{-1} \arctan(\tau_r x)$, where τ_r is the round-trip time for type r-flows; see [29] and [22].

2.3 Generalised constraints: relaxed capacity constraints

A second way in which we generalise the problem is by allowing more general constraint functions than constraining hyperplanes, and where for convenience we move the constraints into the objective function. The bandwidth allocation $\mathbf{x}$ is then defined as the solution of

$$\text{maximise} \quad \sum_{r\in\mathcal{R}} N_r U_r(x_r) - \Gamma\left(N\mathbf{x}; C\right) \tag{9}$$

$$\text{over} \qquad x_r \geq 0, \quad r \in \mathcal{R}, \tag{10}$$

where the strictly concave utility functions U_r are as in (6-8), and Γ is a penalty function, assumed to be convex and non-decreasing in its first argument $N\mathbf{x} := (N_r x_r)_{r\in\mathcal{R}}$. Its second argument represents the notional network link capacities.

This does indeed extend the previous framework, which is recovered by the choice $\Gamma(\mathbf{z}; C) = G_0(\mathbf{z}; C)$, where

$$G_0(\mathbf{z}; C) := \begin{cases} 0 & \text{if } \sum_{r\in\mathcal{R}} A_{jr} z_r \leq C_j, \quad j \in J, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

For instance one might consider continuously differentiable penalty functions $G_\epsilon(\mathbf{z}; C)$ that satisfy

$$\lim_{\epsilon\to 0} G_\epsilon(\mathbf{z}; C) = G_0(\mathbf{z}; C),$$

and study the corresponding allocations $x_r(\epsilon)$.

In addition to convexity, we shall make the following two critical assumptions about the penalty function $\Gamma(\mathbf{z}; C)$.

$$L\Gamma(\mathbf{z}; C) \equiv \Gamma(L\mathbf{z}; LC), \quad \mathbf{z} \in \mathbb{R}_+^R, \quad C \in \mathbb{R}_+^J, L > 0. \tag{12}$$

This condition ensures that the rate allocation x is left unchanged after simultaneously rescaling by some number L both the numbers of flows and the capacities.

Another useful instance of this general framework is as follows. The penalty function Γ can be of the form

$$\Gamma(\mathbf{z}; C) = \sum_{j\in J} \Gamma_j \left(\sum_{r\in\mathcal{R}} A_{jr} z_r; C_j \right), \tag{13}$$

where Γ_j is then a penalty function associated with capacity C_j and is a function of the load on the link, so that in the new optimisation problem the sharp capacity constraint C_j is relaxed. This formulation arises naturally from packet level models, with x_r the mean rate of a stochastic packet generation process. For example, if the resources j correspond to output ports of routers, then there is a limited amount of buffering available, and packets will be dropped if the capacity is exceeded, or more generally marked according to some active queue management technique such as RED [18]. We may interpret $p_j(y_j/C_j)$ as the probability of dropping (or marking) a packet

at resource j when the load on the resource is y_j and its capacity is C_j. In other words when the load on a resource is y_j, a proportion of the load $p_j(y_j/C_j)y_j$ is dropped or marked, and $\Gamma_j(y_j; C_j) = \int_0^{y_j} p_j(\eta/C_j)d\eta$ is the rate at which 'cost' is incurred at the resource. Note that for such Γ, the scaling condition (12) holds.

For example we may consider,

- bufferless resources: we can take $p_j(y_j/C_j) = [y_j - C_j]^+ / [y_j]$;
- small buffers of size b, where small means $b = o(C)$. For example if packets are dropped (or marked) when the buffer content exceeds b then we can use $p_j(y_j/C_j) = \min(1, (y_j/C_j)^b)$. Note that more sophisticated marking strategies such as Virtual Queue marking may produce marking functions of the form $p_j(y_j/C_j) = \min(1, (b+1)(y_j/C_j)^b)$. We may use more accurate models to model the queuing behaviour as a Markov chain in equilibrium - see for example [31].
- moderate buffers, of size $O(\sqrt{C})$. We can use large deviation approximates to derive bounds when the resource is not overloaded, and which behaves in overload like a bufferless resource. See [37] for some more details and interesting discussion on the impact of buffer sizes.
- average delay: a simple choice here is to take $p_j(y_j/C_j)$ proportional to $1/(1-y_j/C_j)$ when $y_j < C_j$, and $+\infty$ otherwise, which acts as a smooth relaxation of a sharp capacity bound.

2.4 Multi-path forwarding

Another instance of the framework (9-10) aims at modeling multi-path traffic forwarding, and is important enough to deserve a separate treatment. Consider the situation where there is a set $\mathcal{S}$ of routes $s \in \mathcal{S}$ available in the network, and each type-r flow may split its traffic among a subset of routes. Let B be the route-flow incidence matrix, with $B_{sr} = 1$ if type r-flows may use route s, and $B_{sr} = 0$ otherwise, where now we let $A = (A_{js})$ denote the link-route incidence matrix.

In this context we define the candidate rate allocations x_r as the solution to the following optimisation problem:

$$\text{Maximise} \quad \sum_{r\in\mathcal{R}} N_r U_r \left(\sum_{s\in\mathcal{S}} B_{sr} x_{sr} \right) - \sum_{j\in J} \Gamma_j \left(\sum_{r\in\mathcal{R}} N_r \sum_{s\in\mathcal{S}} A_{js} B_{sr} x_{sr}; C_j \right) \tag{14}$$

$$\text{over} \quad x_{sr} \geq 0,\ r \in \mathcal{R},\ s \in \mathcal{S}. \tag{15}$$

The variable x_{sr} represents the sending rate of type r-users along route s, and $\Gamma_j(y_j; C_j)$ is the cost for sending at rate y_j over link j, assumed to be convex and non-decreasing.

Alternatively, by introducing the variables $x_r = \sum_{s\in\mathcal{S}} B_{sr} x_{sr}$, and optimising first over the individual route rates x_{sr}, with total rates x_r kept fixed, we see that the per-user rates x_r can also

be characterised as solutions of the generic optimisation problem (9-10), with the specific choice of a cost function:

$$\Gamma(\mathbf{z}; C) = \inf \left\{ \sum_{j \in J} \Gamma_j \left(y_j; C_j \right) \right\}, \tag{16}$$

where the infimum is taken over the set of variables y_j such that

$$\exists y_{sr} \geq 0, \quad \sum_{s \in \mathcal{S}} B_{sr} y_{sr} = z_r, \quad \sum_{s \in \mathcal{S}} A_{js} \sum_{r \in \mathcal{R}} B_{sr} y_{sr} = y_j. \tag{17}$$

The multipath formulation was first described in [24]. Recent work of Han et al. [20] and Kelly and Voice [25] have shown how to construct distributed rate control algorithms which perform this optimisation implicitly. In particular, Kelly and Voice describe per-route controllers and corresponding gain parameter selection strategies which ensure non-oscillatory convergence to the desired allocation, while having the appealing property that each gain is chosen solely on the basis of the round-trip delay of the corresponding route.

2.5 Parallel routing

The multipath forwarding problem we just described assumes that a type r-flow can coordinate responses across routes and hence balance flows across available routes. Although such coordination across routes is conceptually simple, there are issues and problems in implementing it. For example, it breaks the semantics of most current transport layer protocols in the Internet (such as TCP) and hence requires application layer implementation, or use of protocols such as SCTP [41]. An alternative approach, which requires no coordination or balancing across routes comprises setting up for each flow independent connections in parallel, and individually adjusting their flows to maximise their per-route utility, the volume of each type r flow being spread across routes. This corresponds to the following optimisation

$$\text{Maximise} \quad \sum_{r \in \mathcal{R}} N_r \sum_{s \in \mathcal{S}} U_{sr}(x_{sr}) - \sum_{j \in J} \Gamma_j \left(\sum_{r \in \mathcal{R}} N_r \sum_{s \in \mathcal{S}} A_{js} B_{sr} x_{sr}; C_j \right) \tag{18}$$

$$\text{over} \quad x_{sr} \geq 0, \quad s \in \mathcal{S}, \quad r \in \mathcal{R}, \tag{19}$$

where as before B is the flow-route incidence matrix and $A = (A_{js})$ the link-route incidence matrix. As a specific example, parallel TCP connections would correspond to different utility functions in the case where routes have different round trip times, motivating the dependence of U upon both s and r.

3 Flow level stochastic model

We now describe our model of how flows arrive and depart. Our aim is to generalise the stochastic model for file transfers introduced in [32] to include streaming flows. For ease of exposition and in

this section only, we assume the baseline single-route problem, where flows of type r are associated with unique routes, allowing us to use type and route interchangeably. The extensions to multipath or parallel connections are natural and obvious.

Let N_r be the number of document transfers of type r, and let M_r be the number of streaming flows on route r. Define the indicator function $I[r=s]=1$ if $r=s$, $I[r=s]=0$ otherwise. Let $T_sN=(N_r+I[r=s], r\in\mathcal{R})$, with inverse $T_s^{-1}N=(N_r-I[r=s], r\in\mathcal{R})$. We suppose that $(N,M)=(N_r, r\in R; M_r, r\in\mathcal{R})$ is a Markov process, with state space $\mathbb{Z}_+^{\mathcal{R}}\times\mathbb{Z}_+^{\mathcal{R}}$ and non-trivial transition rates

$$q((N,M),(T_rN,M))=\nu_r,\quad q((N,M),(T_r^{-1}N,M))=\mu_rN_rx_r(N+M),\quad r\in\mathcal{R}$$

$$q((N,M),(N,T_rM))=\kappa_r,\quad q((N,M),(N,T_r^{-1}M))=M_r\eta_r,\quad r\in\mathcal{R}$$

for $(N,M)\in\mathbb{Z}_+^{\mathcal{R}}\times\mathbb{Z}_+^{\mathcal{R}}$, where $\mathbf{x}(N)$ is a solution to the optimisation problem (9–10). This corresponds to a model where new file transfers arrive on route r as a Poisson process of rate ν_r, new streaming flows arrive on route r as a Poisson process of rate κ_r, and $x_r(N+M)$ is the bandwidth allocated to each flow on route r, whether it is a file transfer or streaming flow. A file transfer on route r transports a file whose size is exponentially distributed with parameter μ_r, and a streaming flow on route r has an exponentially distributed holding time with parameter η_r.

If $\kappa_r=0, r\in\mathcal{R}$, and in the particular case where the rate allocations are defined via (1–3), then this model reduces to the model introduced by Massoulié and Roberts [32], in which there are no streaming flows, only file transfers. For this case, De Veciana, Lee and Konstantopoulos [13] and Bonald and Massoulié [6] have shown that a sufficient condition for the Markov chain $(N(t), t\geq 0)$ to be positive recurrent is that

$$\sum_r A_{jr}\rho_r<C_j,\quad j\in J, \tag{20}$$

where $\rho_r=\nu_r/\mu_r$; this condition is also necessary [26]. The condition is natural: ρ_r is the load on route r, and we can identify the ratio of the two sides of the inequality (20) as the *traffic intensity* at resource j. Kelly and Williams [26] have explored the behaviour of a fluid model for this case in heavy traffic, when the inequalities (20) are close to being tight, which is a key step towards proving state space collapse. The papers [6, 13, 26] all make use of a fluid model of the Markov process, an approach which we shall adopt for our analysis of the extended model.

In the more general framework (9–10), the natural extension of Condition (20) is the following. For some $\delta\in(0,\infty)^{\mathcal{R}}$,

$$U_r'(\delta_r)>\Gamma_r'(\rho+\delta;C),\quad r\in\mathcal{R}, \tag{21}$$

for all vectors $\Gamma'(\rho+\delta;C)$ that are subgradients of the function Γ in its first argument. When the function Γ is differentiable, there is only one subgradient, which coincides with the ordinary gradient, that is the vector of partial derivatives. At a point where Γ fails to be differentiable, several subgradients may exist; we refer the reader to [38], p.214 for a definition and basic properties of

sub-gradients of convex functions. This condition ensures that the allocation vector x has positive components, and satisfies

$$U_r'(x_r) = \Gamma_r'(N\mathbf{x}; C), \tag{22}$$

where Γ_r' is the partial derivative or a subgradient of the function $\Gamma(\cdot; C)$ at $N\mathbf{x}$.

We now verify that Condition (21) specializes to (20) in the special case where the penalty function Γ captures sharp capacity constraints, and is given by (11). In this case, the subgradient of Γ at each vector $\mathbf{z}$ such that, for all $j \in J$, $\sum_{r\in\mathcal{R}} A_{jr} z_r < C_j$ is the null vector. Thus, provided for all $r \in \mathcal{R}$, $U_r'(\epsilon)$ is positive for small enough $\epsilon > 0$, and (20) holds, Condition (21) is satisfied. Since any strictly concave non-decreasing functions U_r satisfy the first requirement, then indeed (21) holds whenever (20) does.

Let us interpret Condition (21) in the context of multipath forwarding described in Section 2.4, assuming that the link cost function Γ_j represents a sharp capacity constraint, that is $\Gamma_j(z) = 0$ if $z \leq C_j$, and $+\infty$ otherwise. It is easily seen that, for a vector $\mathbf{z} = \{z_r\}_{r\in\mathcal{R}}$, provided there exist y_{sr} and $y = \{y_j\}_{j\in J}$ such that (17) holds, and $y_j < C_j$ for all $j \in J$, then the unique subgradient of Γ at $\mathbf{z}$ is the null vector. Thus, Condition (21) is satisfied provided the loads ρ_r can be split into route loads ρ_{sr} such that we have the natural constraint

$$\sum_{s\in\mathcal{S}, r\in\mathcal{R}} A_{js} B_{sr} \rho_{sr} < C_j, \quad j \in J.$$

We shall henceforth assume that $\kappa_r > 0, r \in R$, and that condition (21) is satisfied.

4 Large capacity scaling: fluid models

Next we consider a fluid model, which can be thought of as a formal law of large numbers approximation under the scaling

$$(\mathbf{n}, \mathbf{m})(t) = \left(\frac{N_L(t)}{L}, \frac{M_L(t)}{L}\right) \quad L \to \infty,$$

where $(N_L(t), M_L(t))$ is the model of the previous Section but with $C_j, j \in J$, and $\nu_r, \kappa_r, r \in \mathcal{R}$, replaced by $LC_j, j \in J$, and $L\nu_r, L\kappa_r, r \in \mathcal{R}$, respectively. The fluid model is an approximation appropriate for the case where $C_j, j \in J$, and $\nu_r, \kappa_r, r \in \mathcal{R}$, are all large, an important case in applications.

Alternatively, in the absence of streaming flows, the fluid model corresponds to the dynamics of the original Markov process describing the number of file transfers, after simultaneous rescaling of both time and space. Such rescaling is known as a *hydrodynamic*, or fluid scaling. It is important because ergodicity of the original stochastic process can be established by proving stability of the fluid model, an approach popularised by Dai [14].

4.1 Fluid model dynamics

An explicit construction of the Markov processes of interest can be obtained from independent unit rate Poisson processes, $\Xi_{i,r}$, $i \in \{1, \dots, 4\}$, $r \in \mathcal{R}$. The trajectories of N, M can then be represented as solutions to the following equations (see e.g. [9]):

$$\begin{cases} N_r(T) = N_r(0) + \Xi_{1,r}(L\nu_r T) - \Xi_{2,r}\left(\int_0^T \mu_r N_r(t) x_r(N(t) + M(t); LC) dt\right), \\ M_r(T) = M_r(0) + \Xi_{3,r}(L\kappa_r T) - \Xi_{4,r}\left(\int_0^T \eta_r M_r(t) dt\right). \end{cases} \tag{23}$$

The rescaled quantities N_r/L, M_r/L are then expected to satisfy in the limit $L \to \infty$ the following equations:

$$\begin{cases} n_r(T) = n_r(0) + \nu_r T - \mu_r \int_0^T n_r(t) x_r(\mathbf{n}(t) + \mathbf{m}(t); C) dt, \\ m_r(T) = m_r(0) + \kappa_r T - \eta_r \int_0^T m_r(t) dt. \end{cases}$$

where we used the homogeneity property (12) according to which $\mathbf{x}(\mathbf{n}, C) = \mathbf{x}(L\mathbf{n}, LC)$. We do not attempt here to prove rigorously the convergence of the rescaled processes to the solutions of these systems of equations, but rather address the reader to [30], for a general reference where conditions under which this type of convergence is valid can be found, or to the forthcoming paper [28], for a full treatment of the single resource case.

In the remainder of this section, we thus focus on the following system of differential equations:

$$\frac{d}{dt} n_r(t) = \nu_r - \mu_r n_r(t) x_r(\mathbf{n}(t) + \mathbf{m}(t); C), \quad r \in \mathcal{R} \tag{24}$$

$$\frac{d}{dt} m_r(t) = \kappa_r - \eta_r m_r(t), \quad r \in \mathcal{R}. \tag{25}$$

Note that our assumption that $\kappa_r > 0, r \in \mathcal{R}$, implies that $m_r(t) > 0, r \in \mathcal{R}, t > 0$.

4.2 Stationary points

Let us first describe the invariant points under the generalised sharing criterion (9-10). We have:

Proposition 1. *Provided the condition (21) is satisfied, the differential equations (24,25) have a unique invariant point,* $(\hat{\mathbf{n}}, \hat{\mathbf{m}})$. *It takes the form*

$$\hat{m}_r = \kappa_r/\eta_r, \quad \hat{n}_r = \rho_r/\hat{x}_r, \quad r \in \mathcal{R}, \tag{26}$$

where the equilibrium allocation $\hat{\mathbf{x}}$ *is the unique solution of*

$$\textit{maximise} \quad \sum_{r \in R} \hat{m}_r U_r(x_r) - \Gamma\left(\hat{\mathbf{m}}\mathbf{x} + \rho; C\right) \tag{27}$$

$$\textit{over} \qquad x_r \geq 0, \quad r \in \mathcal{R}, \tag{28}$$

Proof. The expressions (26) are readily derived from the differential equations (24,25). At any time t, the allocation vector $\mathbf{x}$ is characterised by

$$\left\{U_r'(x_r)\right\}_{r\in\mathcal{R}} = \left\{\Gamma_r'((\mathbf{n}+\mathbf{m})\mathbf{x}; C) - \beta_r\right\}_{r\in\mathcal{R}},$$

where β_r is the Lagrange multiplier associated with the constraint $x_r \geq 0$, and as such satisfies $\beta_r \geq 0$, $\beta_r x_r = 0$, and $\{\Gamma_r'((\mathbf{n}+\mathbf{m})\mathbf{x}; C)\}_{r\in\mathcal{R}}$ is a subgradient of Γ. Therefore at an invariant point we have

$$U_r'(\hat{x}_r) = \Gamma_r'(\rho + \hat{\mathbf{m}}\hat{\mathbf{x}}; C) - \beta_r, \quad r \in \mathcal{R}. \tag{29}$$

This is enough to characterise $\hat{x}_r$ as the solution to (27-28), which we know is unique by strict convexity of the U_r. The stability condition (21) now guarantees that necessarily, $\hat{x}_r > 0$ for all r, and thus $\hat{n}_r$ is finite. □

The invariant point has the following interpretation: the file transfers of type r contribute an irreducible load ρ_r on each resouce they are associated with. The streaming traffic then shares out what remains after the load has been accounted for (obtained via equation (29)) which determines the rate that streams of type r receive and hence under our sharing assumptions also determine the rate that file transfers of type r receive. If we can find *reduced capacities* $\tilde{C}_j$ such that

$$\Gamma_r'(\rho + \hat{\mathbf{m}}\hat{\mathbf{x}}; C) = \Gamma_r'(\hat{\mathbf{m}}\hat{\mathbf{x}}; \tilde{C}) \quad r \in \mathcal{R} \tag{30}$$

then at the invariant point, the file transfers determine the reduced capacities, and the streaming traffic shares the reduced capacity network *as if* it were the only load on this reduced network; the associated rates the streaming traffic receive then determine the rates the file-transfers receive. As remarked above, it is often the case that Γ_j' is a function of the 'load', in which case there is a natural 'reduced capacity'. This is illustrated next in the context of sharp capacity constraints.

4.2.1 Sharp capacity constraints

We next specialise this result to particular cases of interest. Consider first (w, α)-fair bandwidth sharing with sharp capacity constraints (1-3). Define the *reduced capacities*

$$\tilde{C}_j = C_j - \sum_{r\in\mathcal{R}} A_{jr}\rho_r, \quad j \in J. \tag{31}$$

Then the reduced capacity $\tilde{C}_j$ on resource j is just the amount by which inequality (20) fails to be tight. The reduced capacities will determine the capacity available to streaming flows in a sense that we shall now make precise.

Proposition 2. *Provided the condition (20) is satisfied, the differential equations (24,25) have a unique invariant point,* $(\hat{\mathbf{n}}, \hat{\mathbf{m}})$*, given by (26). The equilibrium allocation* $\hat{\mathbf{x}}$ *satisfies*

$$\hat{x}_r = \left(\frac{w_r}{\sum_j p_j A_{jr}}\right)^{1/\alpha}, \tag{32}$$

for some $\mathbf{p} \in \mathbb{R}_+^J$. The pair $(\mathbf{x}, \mathbf{p})$ forms a solution of equation (32) and the conditions

$$p_j \geq 0, \quad p_j \left(\tilde{C}_j - \sum_r A_{jr} \hat{m}_r x_r \right) = 0 \quad j \in J, \tag{33}$$

and together these relations determine $\mathbf{x}$ uniquely.

Proof. The equilibrium rate vector $\hat{\mathbf{x}}$ solves the optimisation problem (27,28), where Γ is now given by (11). Note that for the penalty function G_0, $G_0(\hat{\mathbf{m}}\mathbf{x} + \rho, C) = G_0(\hat{\mathbf{m}}\mathbf{x}, \tilde{C})$, hence it follows that the equilibrium rate vector $\hat{\mathbf{x}}$ may be characterised as the vector of (w, α)-fair allocations of the residual capacities $\tilde{C}$ when there are $\hat{m}_r$ flows along route r. The corresponding characterisation (4,5) then yields (32,33). □

Equations (26,32) describe the vector $\hat{\mathbf{n}}$, of dimension $|\mathcal{R}|$, in terms of $\mathbf{p}$, a vector which may have a much smaller dimension, $|J|$, a phenomenon first noted in the balanced fluid model of [26].

The reduced capacities $(\tilde{C}_j, j \in J)$ that remain after this load is satisfied are available to be shared amongst streaming traffic, and determine the bandwidth allocation to flows on route r for both types of traffic.

When $\kappa_r = 0, r \in R$, the unique invariant point of the fluid model is $\hat{\mathbf{n}} = 0$ [13, 6]. It is notable that the inclusion of streaming traffic within the fluid model forces the components of $\hat{\mathbf{n}}$ to be positive.

We next describe the equilibrium points resulting from the generalised sharing criterion (9,10), when the penalty function Γ is given by (16,17).

4.2.2 Multipath forwarding

Using the notation of Section 2.4, we have at the equilibrium point that

$$B_{sr} > 0 \Rightarrow \quad U_r'(\sum_{s' \in \mathcal{S}} \hat{x}_{s'r}) = \sum_{j \in J} A_{js} \Gamma_j'(\hat{y}_j; C_j) - \beta_{sr} \; r \in R, \; s \in \mathcal{S} \tag{34}$$

where

$$\hat{y}_j = \sum_{s \in \mathcal{S}, r \in \mathcal{R}} A_{js} B_{sr} (\hat{n}_r + \hat{m}_r) \hat{x}_{sr},$$

and β_{sr} is the Lagrange multiplier associated with the constraint $x_{sr} \geq 0$ and satisfies the constraint qualification conditions $\beta_{rs} \geq 0$, $\beta_{rs} \hat{x}_{sr} = 0$, and Γ_j' denotes a subgradient of Γ_j. For any fixed r, it then follows that there exists a critical value p_r such that the "prices" $\sum_{j \in J} A_{js} \Gamma_j'(\hat{y}_j; C_j)$ on any route s such that $B_{sr} = 1$ must coincide with p_r if $\hat{x}_{sr} > 0$, and be less than p_r otherwise.

Denote by ρ_{sr} the fraction $\hat{n}_r \hat{x}_{sr}$ of load ρ_r offered by type r-flows that, in equilibrium, is carried along route s. The above property justifies the following interpretation. With multipath

routing, in equilibrium the load fractions ρ_{sr} are such that the overall cost

$$\sum_{j\in J}\Gamma_j\left(\sum_{s\in\mathcal{S},r\in\mathcal{R}}A_{js}B_{sr}(\rho_{sr}+\hat{m}_r\hat{x}_{sr});C_j\right)$$

is minimised. When no streaming traffic is present, this can be rephrased as follows. Independently of the choice of flow utility functions U_r, under multipath routing, at equilibrium the offered load is split optimally across availables routes.

Even though $\hat{x}_r > 0$, note that it is perfectly possible to have $\hat{x}_{sr} = 0$ for some s– indeed the rates are zero on all 'high-cost' routes with prices strictly larger than p_r.

4.2.3 Parallel routing

In the parallel routing case, at the invariant point

$$U_r'(\hat{x}_{sr}) = \sum_{j\in J}A_{js}B_{sr}\Gamma_j'(\hat{y}_j;C_j) - \beta_{sr}\ r\in\mathcal{R},\ s\in\mathcal{S} \tag{35}$$

hence potentially $\hat{x}_{sr} > 0$ for all s such that $A_{js}B_{sr} = 1$. In other words, the load is spread across all routes that type r traffic can use in a possibly inefficient way; this is discussed in greater detail in Section 5.

4.3 Asymptotic stability

We now establish convergence to the equilibrium point of the dynamics (24,25), assuming the stability condition (21) is satisfied.

In order to do so, we shall first treat the case where there are no streaming flows. The fluid dynamics for the file transfers are then described as

$$\frac{d}{dt}n_r(t) = \nu_r - \mu_r n_r x_r(\mathbf{n}(t);C), \quad r\in\mathcal{R}, \tag{36}$$

where as before $x(\mathbf{n};C)$ solves (9,10). We then have the following result:

Theorem 1. *Under the stability conditions (21), and provided that the penalty function Γ is strictly increasing in each of its coordinates, the function $L(\mathbf{n})$ defined by*

$$L(\mathbf{n}) = \sum_{r\in\mathcal{R}}\frac{1}{\mu_r}\left\{f_r(n_r) - n_r\Gamma_r'(\rho)\right\}, \tag{37}$$

where $\Gamma'(\rho)$ is a sub-gradient of Γ at ρ, and

$$f_r(n) = \int_0^n U_r'\left(\frac{\rho_r}{x}\right)dx \tag{38}$$

is a Lyapunov function for the dynamics (36). These dynamics converge to the set of vectors $\mathbf{n}$ *such that* $\mathbf{n}x(\mathbf{n}; C) = \rho$*, which is in turn also characterised as the set of vectors satisfying*

$$\hat{n}_r = \frac{\rho_r}{U_r'^{-1}\left(\Gamma_r'(\rho)\right)}, \tag{39}$$

where $\Gamma'(\rho)$ *spans the set of sub-gradients of* Γ *at* ρ*. This set of limit points is thus reduced to a single point if and only if* Γ *admits only one sub-gradient at* ρ*.*

Proof. Since we have assumed that Γ is strictly increasing in each of its coordinates, it follows that the rate $x_r(\mathbf{n})$ goes to zero as n_r goes to zero, and hence the trajectories n_r stay away from the boundary of the orthant $\mathbb{R}_+^{\mathcal{R}}$. Define the function ϕ as

$$\phi(\mathbf{z}) := \sum_{r \in \mathcal{R}} n_r U_r\left(\frac{z_r}{n_r}\right) - \Gamma(\mathbf{z}).$$

Since the function ϕ is strictly concave on its domain (that is, the set of points where it is finite)*, and since by the stability condition (21), the vector ρ belongs to its domain, it holds that for any super-gradient $\phi'(\rho)$ of ϕ,

$$\sum_r \phi_r'(\rho)\left(\rho_r - n_r x_r(\mathbf{n})\right) \leq 0,$$

and this inequality is strict unless $\mathbf{nx} = \rho$. For the specific super-gradient $\phi_r'(\rho) = U_r'(\rho_r/n_r) - \Gamma_r'(\rho)$, where $\Gamma'(\rho)$ is the specific sub-gradient of Γ used in the definition of the function L, the left-hand side reads

$$\sum_r \left\{U_r'\left(\frac{\rho_r}{n_r}\right) - \Gamma_r'(\rho)\right\}\left(\rho_r - n_r x_r(\mathbf{n})\right),$$

and is thus equal to

$$\sum_r \frac{\partial L}{\partial n_r}(\mathbf{n}(t))\,\frac{d}{dt}n_r(t) = \frac{d}{dt}L(\mathbf{n}(t)).$$

Thus the value of $L(\mathbf{n})$ decreases strictly along the trajectories of the system, except at points $\mathbf{n}$ such that $\mathbf{nx}(\mathbf{n}; C) = \rho$. Such points are indeed alternatively characterised as solutions of (39). Finally, the function L is such that the level sets $\{\mathbf{n} : L(\mathbf{n}) \leq A\}$ are bounded for all finite A, by stability condition (21). It is thus a proper Lyapunov function. □

We now apply this result to establish stability of the dynamics (24–25).

Corollary 1. *Under the stability condition (21), the dynamics (24–25) are asymptotically stable.*

Proof. We shall only treat the special case where the m_r have already converged to their equilibrium values, $\hat{m}_r$. As the convergence of $\mathbf{m}(t)$ to $\hat{\mathbf{m}}$ does not depend on the evolution of $\mathbf{n}(t)$, the general case can be deduced by continuity arguments. We now show that the n_r evolve according

*Strict concavity of ϕ follows from concavity of the two terms in its definition, and strict concavity of its first term.

to (36) for some suitable choice of a penalty function $\tilde{\Gamma}$. Indeed, (36) holds, with the rate vector $\mathbf{x}$ solving

$$\begin{aligned} \text{maximise} \quad & \phi(\mathbf{x},\mathbf{y}) := \sum_{r\in R} n_r U_r(x_r) + \hat{m}_r U_r(y_r) - \Gamma\left(\mathbf{n}\mathbf{x} + \hat{\mathbf{m}}\mathbf{y}\right) \\ \text{over} \quad & x_r, y_r \geq 0,\ r \in R. \end{aligned}$$

Performing the optimisation over the y_r first, the corresponding allocation vector $\mathbf{x}$ is again the solution of (9–10), with Γ replaced by $\tilde{\Gamma}$, which is defined by

$$\begin{aligned} \tilde{\Gamma}(\mathbf{z}) \quad := \inf \quad & \left\{ \Gamma(\mathbf{z} + \hat{\mathbf{m}}\mathbf{y}) - \sum_{r\in R} \hat{m}_r U_r(y_r) \right\}, \\ \text{over} \quad & y_r \geq 0, r \in R. \end{aligned} \tag{40}$$

It is readily seen that $\tilde{\Gamma}$ is increasing in each coordinate since each U_r is assumed to be strictly increasing. Convexity of $\tilde{\Gamma}$ also holds: this can be verified directly, but also follows from recognising that $\tilde{\Gamma}$ is the inf-convolution of two convex functions, and as such convex itself. □

Remark 1. *By comparing equations (26-28) of Proposition 1 with (39), we obtain the following identification:* $\tilde{\Gamma}'(\rho; C) = \Gamma'(\rho + \hat{\mathbf{n}}\hat{\mathbf{x}}; C)$, *where* $\tilde{\Gamma}$ *is as in (40).*

Besides, the proof of Corollary 1 suggests the following interpretation. The impact of streaming flows on file transfers can be simply captured by a suitable change in the penalty function, namely replacing the original function Γ *by* $\tilde{\Gamma}$ *as defined in (40).*

5 Uncoordinated parallel versus balanced multipath routing

In the present section we compare the performance of coordinated multipath routing with parallel routing. For ease of exposition we assume only file transfers are present. This involves no loss of generality, since we could capture the influence of streaming traffic by redefining the cost function, as in Remark 1 above.

The stability results of Theorem 1 provide stability conditions for multipath forwarding, as this fits in the general allocation framework (9-10). However we have not provided general stability results for parallel connections. 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

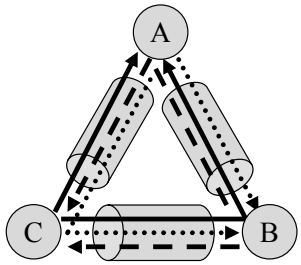

Figure 1: Example network where parallel uncoordinated connections lead to inefficiency

are allowed, the capacity region is described by

$$\begin{cases} \rho_B + \rho_C \leq 2, \\ \rho_C + \rho_A \leq 2, \\ \rho_A + \rho_B \leq 2. \end{cases} \tag{41}$$

In the symmetric case when all three offered loads coincide, the stability condition reads $\rho \leq 1$. 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 (resp. n_B, n_C) denote the number of type-A (respectively, $-B$, $-C$) 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

$$\begin{cases} n_A \left(\frac{1}{p_A}\right)^{1/\alpha} + n_B \left(\frac{1}{p_A+p_C}\right)^{1/\alpha} + n_C \left(\frac{1}{p_A+p_B}\right)^{1/\alpha} = 1, \\ n_B \left(\frac{1}{p_B}\right)^{1/\alpha} + n_C \left(\frac{1}{p_A+p_B}\right)^{1/\alpha} + n_A \left(\frac{1}{p_B+p_C}\right)^{1/\alpha} = 1, \\ n_C \left(\frac{1}{p_C}\right)^{1/\alpha} + n_A \left(\frac{1}{p_B+p_C}\right)^{1/\alpha} + n_B \left(\frac{1}{p_A+p_C}\right)^{1/\alpha} = 1. \end{cases} \tag{42}$$

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

$$\frac{d}{dt} n_i(t) = \nu - \mu x_i(\mathbf{n}(t)), \quad i = A, B, C. \tag{43}$$

We have the following:

Proposition 3. *Define ρ as λ/μ, and*

$$\rho^* := \frac{1+2^{-1/\alpha}}{1+2^{1-1/\alpha}}. \tag{44}$$

The solution $\mathbf{n}(t)$ to the system of differential equations (43) 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 $\rho < \rho^*$.

Proof: We shall rely on monotonicity properties of the allocations, which we summarise in the following lemma, the proof of which relies on elementary manipulations of (42), and is left as an exercise.

Lemma 1. *Assume $n_A \le n_B \le n_C$. Then it holds that:*

$$p_A \le p_B \le p_C, \tag{45}$$

$$n_A \left(\frac{1}{p_A}\right)^{1/\alpha} \le n_B \left(\frac{1}{p_B}\right)^{1/\alpha} \le n_C \left(\frac{1}{p_C}\right)^{1/\alpha}, \tag{46}$$

$$n_A \left(\frac{1}{p_B+p_C}\right)^{1/\alpha} \le n_B \left(\frac{1}{p_A+p_C}\right)^{1/\alpha} \le n_C \left(\frac{1}{p_A+p_B}\right)^{1/\alpha}. \tag{47}$$

We shall now deduce the following. Suppose $n_A \le n_B \le n_C$. Then it holds that

$$n_A x_A \le \rho^*, \tag{48}$$

where ρ^* is defined in (44). Indeed, assume that it is not so, that is

$$n_A \left(\frac{1}{p_A}\right)^{1/\alpha} + n_A \left(\frac{1}{p_B+p_C}\right)^{1/\alpha} > \rho^*. \tag{49}$$

Then, using the inequalities (45), we obtain that necessarily

$$n_A \left(\frac{1}{p_A}\right)^{1/\alpha} \left(1+2^{-1/\alpha}\right) > \rho^*. \tag{50}$$

On the other hand, (49) and the first equation of (42) imply that

$$\rho^* - n_A \left(\frac{1}{p_B+p_C}\right)^{1/\alpha} + n_B \left(\frac{1}{p_A+p_C}\right)^{1/\alpha} + n_C \left(\frac{1}{p_A+p_B}\right)^{1/\alpha} < 1.$$

The sum of the two middle terms in the left-hand side is positive, by (47), so that

$$n_C \left(\frac{1}{p_A+p_B}\right)^{1/\alpha} < 1-\rho^*.$$

By (45), this further implies

$$n_C \left(\frac{1}{p_C}\right)^{1/\alpha} < 2^{1/\alpha}(1-\rho^*).$$

In view of (46), this implies

$$n_A \left(\frac{1}{p_A}\right)^{1/\alpha} < 2^{1/\alpha}(1-\rho^*) = \frac{\rho^*}{1+2^{-1/\alpha}},$$

which contradicts (50), thus establishing the desired inequality (48).

Thus, when $n_A \le n_B \le n_C$, we necessarily have that

$$\frac{d}{dt} n_A(t) \ge \mu\left(\rho - \rho^*\right).$$

Thus the minimum of the three components increases at rate $\mu(\rho - \rho^*)$, which establishes the first half of the proposition.

The second half is established in a similar manner: by a direct adaptation of the argument used to establish (48) one can show that, when $n_A \le n_B \le n_C$, the allocation x_C is at least ρ^*. Thus, the largest of n_A, n_B and n_C decreases to zero at speed at least $\mu(\rho^* - \rho)$, which establishes the second half of the proposition. □

Remark 2. *In the case where $\rho < \rho^*$, the second half of the proposition, combined with Dai's stability criterion [14] shows that the original stochastic system is ergodic.*

It can also be shown with additional work that the original stochastic system is transient under the assumption $\rho > \rho^*$. We omit the detailed arguments in the present paper. Together, these results show that ρ^* is indeed the exact capacity of the triangle network of Figure 1 under symmetric loads.

As we have just seen, the use of parallel, uncoordinated connections can lead to strictly smaller capacity regions than coordinated multipath transfers. We now discuss a specific case of interest where the capacity region is the same for coordinated multipath and for uncoordinated parallel connections.

Consider the case where each file transfer type r can use several paths $p \in \mathcal{P}_r$, 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 consider first 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, can be described in the following simple manner:

$$\sum_{r \in \mathcal{S}} \rho_r < \sum_{\ell \in \mathcal{L}(\mathcal{S})} C_\ell, \quad \mathcal{S} \subset \mathcal{R}, \tag{51}$$

where the subset of links $\mathcal{L}(\mathcal{S})$ is defined to be the union of the sets $\mathcal{P}(r)$ for all $r \in \mathcal{S}$. Indeed, clearly the conditions with non-strict inequalities instead of strict ones are 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

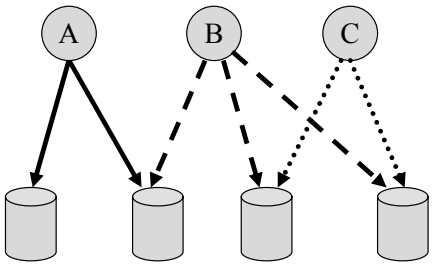

Figure 2: Example network with 1-hop routes: parallel unccordinated connections achieve maximal stability region

are also sufficient is known as Hall's theorem (see e.g. [5], p.77) in the case where 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 usual 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

$$x_r = \sum_{\ell \in \mathcal{P}(r)} x_{r\ell}, \tag{52}$$

where the $x_{r\ell}$ maximise

$$\sum_r n_r \sum_{\ell \in \mathcal{P}(r)} U_{r\ell}(x_{r\ell}) - \sum_{\ell \in \mathcal{L}} \Gamma_\ell \left(\sum_{r:\ell \in \mathcal{P}(r)} n_r x_{r\ell} \right). \tag{53}$$

We now establish the following

Proposition 4. *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 (51), the solutions to the system of differential equations*

$$\frac{d}{dt} n_r(t) = \nu_r - \mu_r n_r x_r \tag{54}$$

where x_r is specified by (52–53) return to zero in time at most $\theta \sum_{r \in \mathcal{R}} n_r(0)$ for some suitable constant θ.

Proof: Let $W(t) = \sum_{r \in \mathcal{S}} \mu_r^{-1} n_r(t)$ denote the expected amount of work present in the system. We argue that, whenever $W(t) > 0$, it holds that

$$\frac{d}{dt} W(t) \leq -\epsilon,$$

where

$$\epsilon = \min_{\mathcal{S}\subset\mathcal{R},\mathcal{S}\neq\emptyset} \left[\sum_{\ell\in\mathcal{L}(\mathcal{S})} C_\ell - \sum_{r\in\mathcal{S}} \rho_r \right].$$

Indeed, let $\mathcal{R}_1(t)$ denote the set of user types r such that $n_r(t) > 0$. By the assumption of boundedness of the ratio of derivatives of utility functions, it holds that the capacity C_ℓ of links ℓ in $\mathcal{L}(\mathcal{R}_1)$ is entirely used by users of types r belonging to $\mathcal{R}_1$. Thus, it holds that

$$\begin{aligned} W(t+h) - W(t) &= \textstyle\sum_{r\in\mathcal{R}} \int_t^{t+h} \mathbf{1}_{r\in\mathcal{R}_1(u)} \left(\rho_r - n_r x_r(\mathbf{n}(u))\right) du \\ &= \textstyle\sum_{\mathcal{S}\subset\mathcal{R}} \int_t^{t+h} \mathbf{1}_{\mathcal{R}_1(u)=\mathcal{S}} \left(\sum_{r\in\mathcal{S}} \rho_r - \sum_{\ell\in\mathcal{L}(\mathcal{S})} C_\ell \right) du \\ &\leq -\epsilon \textstyle\int_t^{t+h} \mathbf{1}_{\mathcal{R}_1(u)\neq\emptyset} du. \end{aligned}$$

This establishes that $W(t)$ decreases indeed at rate at least ϵ until it reaches 0, concluding the proof. □

Remark 3. *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 the comments in Section 4.2.2 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:

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

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

$$\textit{Maximise} \quad n\left[U_1(x_1) + U_2(x_2)\right] - \Gamma_1(n x_1) - \Gamma_2(n x_2).$$

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.

6 Second-order properties

The aim of the present section is to establish diffusion approximations for the rescaled Markov processes $(L^{-1}N_r(t), L^{-1}M_r(t))_{r\in\mathcal{R}}$ of Section 3 around the fluid limits $n_r(t), m_r(t)$, evolving

according to (24,25), that have been studied in Section 4. The derivations in this section are purely formal, and no rigorous justification is provided; we address the reader to [28] for a rigorous treatment of the single link scenario. This section is included for completeness and to illustrate how subtler performance issues could be addressed.

We introduce the perturbation processes

$$\begin{cases} u_r(t) = \frac{1}{\sqrt{L}} \left(N_r(t) - L n_r(t) \right), \\ v_r(t) = \frac{1}{\sqrt{L}} \left(M_r(t) - L m_r(t) \right), \end{cases}$$

together with the noise processes

$$\xi_{i,r}(t) = \frac{1}{\sqrt{L}} \left(\Xi_{i,r}(Lt) - Lt \right), \quad i \in \{1, \cdots, 4\}, \quad r \in R, \quad t > 0,$$

where the $\Xi_{i,r}$ are the unit rate Poisson processes appearing in the representation (23). By taking $\mathbf{n}(0) = \hat{\mathbf{n}}$, and $\mathbf{m}(0) = \hat{\mathbf{m}}$, assuming differentiability of the allocation vector $\mathbf{x}$ with respect to the flow numbers n_r, we obtain formally the limiting equations for the perturbation processes $\mathbf{u}$, $\mathbf{v}$:

$$\begin{aligned} u_r(T) &= u_r(0) + \xi_{1,r}(\nu_r T) - \xi_{2,r}(\nu_r T) \\ &\quad -\mu_r \int_0^T \hat{x}_r u_r(t) dt - \mu_r \int_0^T \hat{n}_r \sum_{s \in R} \left(\frac{\partial x_r}{\partial n_s} \right) (\hat{\mathbf{n}} + \hat{\mathbf{m}}; C) \left(u_s(t) + v_s(t) \right) dt, \\ v_r(T) &= v_r(0) + \xi_{3,r}(\kappa_r T) - \xi_{4,r}(\kappa_r T) - \eta_r \int_0^T v_r(t) dt. \end{aligned}$$

Replacing the noise processes $\xi_{i,r}$ by standard Wiener processes, we can alternatively write these as the following system of stochastic differential equations:

$$d \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = - \begin{pmatrix} \lceil \mu \hat{\mathbf{x}} \rfloor + \lceil \mu \hat{\mathbf{n}} \rfloor D & \lceil \mu \hat{\mathbf{n}} \rfloor D \\ 0 & \lceil \eta \rfloor \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} dt + \begin{pmatrix} \lceil \sqrt{2\nu} \rfloor & 0 \\ 0 & \lceil \sqrt{2\kappa} \rfloor \end{pmatrix} d \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}, \quad (55)$$

where $\lceil a \rfloor$ represents the diagonal matrix with entries a_r on the diagonal, D is the square matrix with entries $\partial x_r / \partial n_s$ evaluated at $\hat{\mathbf{n}} + \hat{\mathbf{m}}$, and W_1, W_2 are independent, standard $|\mathcal{R}|$-dimensional Wiener processes. In the above, for compactness of representation we have also replaced $\xi_{1,r} - \xi_{2,r}$ by $\sqrt{2} W_{1,r}$ where $W_{1,r}$ is again a standard Wiener process, and similarly for $\xi_{3,r} - \xi_{4,r}$.

This characterises the perturbation processes $\mathbf{u}$, $\mathbf{v}$ as coupled Ornstein-Uhlenbeck processes. Correlations can then be determined from the matrices in the above representation. The matrix D is determined as follows, in the special case where Γ and U are twice differentiable. By differentiating equation (22), denoting by H the Hessian matrix of Γ, evaluated at $\hat{\mathbf{n}}\hat{\mathbf{x}}$, one obtains

$$\left\lceil U''(\hat{x}) \right\rfloor D = H \left(\lceil \hat{\mathbf{x}} \rfloor + \lceil \hat{\mathbf{n}} \rfloor D \right),$$

so that

$$D = - \left(H \lceil \hat{\mathbf{n}} \rfloor - \left\lceil U''(\hat{x}) \right\rfloor \right)^{-1} H \lceil \hat{\mathbf{x}} \rfloor. \quad (56)$$

For the sake of illustration, consider the case where Γ is as in (13), where the individual penalty functions Γ_j are twice differentiable. The Hessian matrix H then reads:

$$H = A^T \left\lceil \gamma'' \right\rfloor A$$

where the diagonal entry γ''_j is given by

$$\frac{1}{C_j^2}\Gamma''_j\left(\frac{\sum_r A_{jr}\hat{x}_r(\hat{n}_r+\hat{m}_r)}{C_j}\right).$$

Remark 4. *By removing the m-component, that describes the streaming flows, in the above Ornstein-Uhlenbeck process, which is formally done by setting κ and υ to zero, we obtain a reduced Ornstein-Uhlenbeck process for the fluctuations in the n-component.*

This is counter-intuitive, as one expects, from standard heavy traffic theory, the diffusion approximation of the m-component to behave like a reflected brownian motion instead. Such an expectation is backed by simulation results reported in [27]. However, the above Ornstein-Uhlenbeck process limit is obtained based on the assumption that the penalty function Γ is twice differentiable, whereas reflected Brownian motions in heavy traffic limits arise in the presence of sharp capacity constraints, hence non-differentiable Γ. These issues are investigated in greater detail in [28].

7 Conclusion

We have studied a flow level model of Internet congestion control, that represents the randomly varying number of flows present in a network. Bandwidth was assumed to be dynamically shared between file transfers and streaming traffic, according to a fairness criterion that includes TCP friendliness as a special case. Through the construction of an appropriate Lyapunov function we have established stability, under conditions, for a fluid model of the system. The presence of fair-sharing streaming traffic results in a non-degenerate fluid model. Analysis of the model suggests that file transfers are seen by streaming traffic as reducing the available capacity, whereas for file transfers the presence of streaming traffic amounts to a simple modification in the network penalty function.

While we have assumed that streaming traffic fairly shares the capacity with file transfers, our model can be adapted to the case where streaming flows have a minimum or fixed bandwidth requirement, and admission control is used so that the aggregate rate used by streaming traffic competes fairly with file transfers. For details see [27].

The general bandwidth allocation criterion we have considered encompasses balanced multi-path routing, which could be implemented by modifying the existing TCP transport protocol as in the proposals of [20, 25]. We have also compared the performance of such routing to parallel, uncoordinated routing, which may be implemented with fewer changes to existing protocols. We have shown that the latter may strictly reduce the capacity region of the network as compared to the former. This strengthens the case for deploying modified versions of Internet transport protocols as those described in [20, 25].

Finally, we have formally identified second-order diffusion approximations to the first-order fluid limits of the number of flows in progress. These provide a basis for refined performance evaluation of integrated network-wide data transfer.

References

[1] E. Altman, K.E. Avrachenkov and C. Barakat, TCP network calculus: the case of large delay-bandwidth product, in Proceedings of IEEE Infocom, 2002.

[2] N. Antunes, C. Fricker, F. Guillemin and P. Robert, Integration of streaming services and TCP data transmission in the Internet, In Proceedings Performance, 2005.

[3] A. Bain and P. B. Key, Modelling the performance of distributed admission control for adaptive applications, *Performance Evaluation Review*, December 2001.

[4] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, J. Roberts, Statistical bandwidth sharing: a study of congestion at flow level. In *Proceedings of SIGCOMM 2001*.

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

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

[7] T. Bonald and A. Proutière, On performance bounds for the integration of elastic and adaptive streaming flows, In Proceedings of ACM Sigmetrics / Performance 2004.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[9] P. Brémaud, *Markov chains, Gibbs fields, Monte Carlo simulation, and queues.* Springer-Verlag, New York, 1999.

[10] L. Breslau, E. W. Knightly, S. Shenker, I. Stoica, and H. Zhang, Endpoint admission control: Architectural issues and performance. In *Proceedings of SIGCOMM 2000*, 57–69, 2002.

[11] T. Bu and D. Towsley, Fixed Point Approximation for TCP behavior in an AQM Network. In *Proceedings of ACM SIGMETRICS 2001*.

[12] C. A. Courcoubetis, A. Dimakis, and M. I. Reiman, Providing bandwidth guarantees over a best-effort network: call admission and pricing. *IEEE INFOCOM*, 459–467, 2001.

[13] G. de Veciana, T-J. Lee and T. Konstantopoulos, Stability and performance analysis of networks supporting elastic services, *IEEE/ACM Trans. on Networking* **9**, 2–14, 2001.

[14] J. Dai, On positive recurrence of mutliclass queueing networks: a unified approach via fluid limit models, Ann. of Applied Probability **5**, 49-77, 1995.

[15] F. Delcoigne, A. Proutière and G. Régnié, Modeling integration of streaming and data traffic, Performance Evaluation 55 (3-4), 185-209, 2004.

[16] S. Floyd and K. Fall, Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Transactions on Networking* **7**, 458–472, 1999.

[17] Floyd, S., M. Handley, J. Padhye and J. Widmer (2000), Equation-based congestion control for unicast applications. In *Proc. ACM SIGCOMM 2000*, 43–54, Stockholm, 2000.

[18] S. Floyd and V. Jacobson, Random early detection gateways for congestion avoidance, IEEE/ACM Trans. Networking, **4**, 1993.

[19] R.J. Gibbens, S.K. Sargood, C. Van Eijl, F.P. Kelly, H. Azmoodeh, R.N. Macfadyen and N.W. Macfadyen, Fixed-point models for the end-to-end performance analysis of IP networks. 13th ITC Specialist Seminar: IP Traffic Measurement, Modeling and Management, Monterey, California, 2000.

[20] 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.

[21] IP monitoring project, Sprint labs. `http://ipmon.sprintlabs.com`

[22] F.P. Kelly, Mathematical modeling of the Internet, in "Mathematics Unlimited - 2001 and Beyond" (Editors B. Engquist and W. Schmid). Springer-Verlag, Berlin, 685–702, 2001.

[23] F.P. Kelly, P.B. Key, and S. Zachary, Distributed admission control. *IEEE Journal on Selected Areas in Communications*, **18**, 2617–2628, 2000.

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

[25] 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.

[26] F.P. Kelly and R. J. Williams, Fluid model for a network operating under a fair bandwidth-sharing policy. Annals of Applied Probability **14** 1055–1083, 2004.

[27] P. Key, L. Massoulié, A. Bain and F. Kelly, Fair Internet traffic integration: network flow models and analysis, Annals of Telecommunications, **59**, 1338–1352, 2004.

[28] 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 .

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

[30] T.G. Kurtz, Strong Approximation theorems for density dependent Markov chains, Stochastic Process. Appl. **6** 223–240, 1978.

[31] P. Kuusela, P. Lassila, J. Virtamo and P. Key, Modeling RED with Idealized TCP Sources, Proceedings of IFIP ATM & IP 2001, Budapest, Hungary, 155–166, 2001.

[32] L. Massoulié and J. Roberts, Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems* **15**, 185–201, 2000.

[33] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, The macroscopic behaviour of the TCP congestion avoidance algorithm. *Computer Communication Review* **27**, 67–82, 1997.

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

[35] R. Núñez-Queija, J.L. van den Berg and M.R.H. Mandjes, Performance evaluation of strategies for integration of elastic and stream traffic, In Proceedings ITC-16, 1999.

[36] J. Padhye, V. Firoiu, D. Towsley and J. Kurose, Modeling TCP Reno performance: a simple model and its empirical validation. *IEEE/ACM Transactions on Networking* **8**, 133–145, 2000.

[37] G. Raina and D.J. Wischik, Buffer sizes for large multiplexers: TCP queueing theory and instability analysis, In Proceedings of EuroNGI conference, Rome, April 2005.

[38] T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[39] M. Roughan, A. Erramilli and D. Veitch, Network performance for TCP Networks, Part I: persistent sources. In Proceedings of ITC'17 Brasil, September 2001.

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

[41] Streaming Control Transmission Protocol. http://www.sctp.org