Multipath TCP Algorithms: Theory and Design

Qiuyu Peng
EE, California Institute of Technology
qpeng@caltech.edu

Anwar Walid
Alcatel-Lucent Bell Labs
anwar@research.bell-labs.com

Steven H. Low
CMS & EE, California Institute of Technology
slow@caltech.edu

ABSTRACT

Multi-path TCP (MP-TCP) has the potential to greatly improve application performance by using multiple paths transparently. We propose a fluid model for a large class of MP-TCP algorithms and identify design criteria that guarantee the existence, uniqueness, and stability of system equilibrium. We characterize algorithm parameters for TCP-friendliness and prove an inevitable tradeoff between responsiveness and friendliness. We discuss the implications of these properties on the behavior of existing algorithms and motivate a new design that generalizes existing algorithms. We use ns2 simulations to evaluate the proposed algorithm and illustrate its superior overall performance.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Distributed networks*

Keywords

Multipath TCP, Congestion control.

1. INTRODUCTION

Traditional single-path TCP traverses one route so that only one access interface can be used by the application. This limits performance when there are multiple interfaces/routes available, e.g. most smart phones are enabled with both cellular and WiFi access, and communicating servers in data centers are connected through multiple routes. Multi-path TCP (MP-TCP) has the potential to greatly improve application performance by using multiple paths transparently. The Internet Engineering Task Force (IETF) has started the MP-TCP Working Group [3], which is chartered to develop mechanisms that enable an application to take advantage of available paths. MP-TCP is envisioned to co-exist with single-path TCP such that applications that use MP-TCP can benefit from using available capacity on multiple paths without degrading the performance of applications that use single-path TCP.

Various congestion control algorithms have been proposed as an extension of TCP NewReno for MP-TCP. A straightforward extension is to run TCP NewReno on each subpath, e.g. [5, 6]. This basic extension can lead to a highly unfair bandwidth allocation for single-path TCP users when their paths share bottleneck links with paths used by MP-TCP users. In order to resolve these unfairness issues, the Coupled [1] algorithm [4, 7] is proposed and shown to be fair to single-path TCP. The underlying reason is that Coupled algorithm and TCP NewReno are associated with the same utility function. Recently in [13], it is found that Coupled algorithm can not adapt fast enough in a dynamic network environment, such that it only uses one route even if there are multiple available routes. A different algorithm is proposed in [13] (We refer to this algorithm as the Max algorithm), which is claimed to be both fair to single-path TCP and more responsive than the Coupled algorithm. However, as we show in our simulation, the Max algorithm is still unfair to single-path TCP, and the unfairness is exaggerated when the round trip time of each subpath is different.

To develop improved algorithms, certain questions need to be addressed. First, the design criteria for MP-TCP to converge to a unique equilibrium need to be identified. For single-path TCP algorithm, one can associate a strict concave utility function for each source so that the congestion control algorithm implicitly solves a network utility maximization problem [8, 11]. The utility maximization interpretation provides an intuitive approach for showing the existence and uniqueness of equilibrium. For MP-TCP, it will be shown that the utility maximization interpretation fails to hold in general, necessitating the need to develop new tools for studying the equilibrium of MP-TCP algorithms. Second, the relations among different performance metrics, e.g. fairness, responsiveness and window fluctuation, need to be identified. A theoretical understanding of the design space for MP-TCP is needed so that improved algorithms can be developed. The performance metrics that we focus on are defined as follows:

- *TCP Friendliness*: MP-TCP flows can harm single-path TCP flows by taking more bandwidth when they share bottleneck links. The friendliness measure describes the degree of aggressiveness of MP-TCP flows towards single-path TCP flows.
- *Responsivness*: MP-TCP algorithms should adapt fast in dynamic network environments. Responsiveness characterizes the rate of algorithm convergence.
- *Window Fluctuation*: The congestion window fluctuates due to the additive-increase, multiplicative-decrease (AIMD) property of loss-based TCP.

In this paper, we will address the above questions and propose a new MP-TCP algorithm. The main contributions of this paper are three-fold: First, we show that there is no associated utility function for some MP-TCP algorithms, e.g. the Max algorithm. Thus the intuitive way of showing a

[1]This name comes from [13]

SIGMETRICS'13, June 17-21, 2013, Pittsburgh, PA, USA.

unique stable equilibrium by utility maximization interpretation fails. We consider a unified fluid model that covers a broad class of MP-TCP algorithms and prove, under mild conditions: (i) the existence and uniqueness of equilibrium, (ii) asymptotical convergence of the algorithms. Indeed, algorithms that fail to satisfy the proposed conditions, e.g. the Coupled algorithm, are likely to be unstable and may have multiple equilibria as shown in [13]. Second, we define the performance metrics of TCP friendliness, responsiveness and window fluctuation in the context of MP-TCP, and explore the algorithm design space by characterizing tradeoffs among these performance metrics. Third, based on our understanding of the design space, we propose a new MP-TCP algorithm. We evaluate the proposed algorithm and the existing algorithms using ns2 simulation, and illustrate the superior overall performance of the proposed algorithm.

We now summarize our proposed MP-TCP algorithm. Each source s has a set of routes r. Each route r maintains a congestion window w_r and can measure its round trip time τ_r. The window adaptation are as follows:

- For each ACK on route $r \in s$,

$$w_r \leftarrow w_r + \frac{x_r}{\tau_r \left(\sum x_k\right)^2} \left(\frac{1+\alpha_r}{2}\right) \left(\frac{4+\alpha_r}{5}\right) \tag{1}$$

- For each packet loss on route $r \in s$,

$$w_r \leftarrow \max\left\{w_r \left(1 - \frac{1}{2}\alpha_r\right), 1\right\} \tag{2}$$

where $x_r := w_r/\tau_r$ and $\alpha_r := \frac{\max\{x_k\}}{x_r}$.

The rest of the paper is structured as follows. In Section 2, we develop a unified model for MP-TCP and use it to model existing algorithms. In Section 3 we prove several structural properties, focusing on design criteria that guarantee the existence, uniqueness, and stability of system equilibrium, TCP-friendliness, responsiveness, and inevitable tradeoff among these properties. In section 4, we discuss the implications of these properties and design criteria on the existing algorithms. This motivates our new MP-TCP algorithm and we explain our design rationale. Finally in Section 5 we use ns2 simulation to compare the performance of the proposed algorithm with the existing algorithms. The paper is concluded in section 6.

2. MULTIPATH TCP MODEL

In this section we first propose a fluid model of MP-TCP and then use it to model existing MP-TCP algorithms in the literature. In the next section we will present some structural properties of the model and discuss their implications on the design choices made in these MP-TCP algorithms.

Unless otherwise specified, a capital letter is used to denote a matrix or a set, depending on the context. A matrix P, not necessarily symmetric, is defined as positive definite if $\mathbf{x}^T P \mathbf{x} > 0$ for any $\mathbf{x} \neq \mathbf{0}$. We emphasize that, unlike in conventional use, the definition here does not require P to be symmetric; in fact the matrices of interest in this paper are not symmetric, making the analysis difficult. Given two matrix A, B, $A \succeq B$ means $A - B$ is positive semidefinite. A boldface letter is used to denote a vector. $\|\mathbf{x}\|_n := (\sum x_i^n)^{1/n}$ defines the n norm of a vector $\mathbf{x}$. Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} \geq \mathbf{y}$ means $x_i \geq y_i$ for all components. For a vector $\mathbf{x}$, $diag\{\mathbf{x}\}$ is a diagonal matrix with entries given by $\mathbf{x}$.

2.1 Fluid model

Consider a network that consists of a set $L = \{1, \ldots, L\}$ of links with finite capacities c_l. The network is shared by a set $S = \{1, \ldots, S\}$ of sources. Available to source $s \in S$ is a fixed collection of routes r. A route r consists of a set of links l. We abuse notation and use s both to denote a source and the set of routes r available to it, depending on the context. Likewise, r is used both to denote a route and the set of links l in the route. Let $R := \{r \mid r \in s, s \in S\}$ be the collection of all routes. Let $H \in \{0,1\}^{|L| \times |R|}$ be the routing matrix: $H_{lr} = 1$ if link l is in route r (denoted by '$l \in r$'), and 0 otherwise.

For each route $r \in R$, τ_r denotes its round trip time (RTT). For simplicity we assume τ_r are constants. Each source s maintains a congestion window $w_r(t)$ at time t for every route $r \in s$. Let $x_r(t) := w_r(t)/\tau_r$ represent the throughput on route r. Each link l maintains a congestion price $p_l(t)$ at time t. Let $q_r(t) := \sum_{l \in r} p_l(t)$ be the (approximate) aggregate price on route r. In this paper $p_l(t)$ represents the packet loss probability at link l and $q_r(t)$ represents the packet loss probability on route r.

We associate three state variables $(x_r(t), w_r(t), q_r(t))$ for each route $r \in s$. Let $\mathbf{x}_s(t) := (x_r(t), r \in s)$, $\mathbf{w}_s(t) := (w_r(t), r \in s)$, $\mathbf{q}_s(t) := (q_r(t), r \in s)$. Also let $\tau_s := (\tau_r, r \in s)$. Then $(\mathbf{x}_s(t), \mathbf{w}_s(t), \mathbf{q}_s(t))$ represents the corresponding state variables for each source $s \in S$. For each link l, let $y_l(t) := \sum_{r \in R} H_{lr} x_r(t)$ be its aggregate traffic rate.

Congestion control is a distributed algorithm that adapts $\mathbf{x}(t)$ and $\mathbf{p}(t)$ in a closed loop. Motivated by the AIMD algorithm as in TCP Reno, we model MP-TCP by

$$\dot{x}_r = k_r(\mathbf{x}_s) \left(\phi_r(\mathbf{x}_s) - \frac{1}{2} q_r\right)_{x_r}^+ \qquad r \in s \quad s \in S \tag{3}$$

$$\dot{p}_l = \gamma_l \left(y_l - c_l\right)_{p_l}^+ \qquad l \in L, \tag{4}$$

where $(a)_x^+ = a$ for $x > 0$ and $\max\{0, a\}$ for $x \leq 0$. We omit the time t in the expression for simplicity. Eqn. (3) models how sending rates are adapted in the congestion avoidance phase of TCP at each end system and (4) models how the congestion price is updated at each link by AQM. The MP-TCP algorithm installed at source s is specified by (K_s, Φ_s), where $K_s(\mathbf{x}_s) := (k_r(\mathbf{x}_s), r \in s)$ and $\Phi_s(\mathbf{x}_s) := (\phi_r(\mathbf{x}_s), r \in s)$. Here $K_s(\mathbf{x}_s) \geq 0$ is a positive gain that determines the dynamic property of the algorithm. $\Phi_s(\mathbf{x}_s)$ determines the equilibrium properties of the algorithm. The AQM algorithm is specified by γ_l, where $\gamma_l > 0$ is a positive gain that determines the dynamic property. This is a simplified model for the RED algorithm that assumes the loss probability is proportional to the backlog, and is used in, e.g., [7, 10, 11].

2.2 Existing MP-TCP algorithms

We first show how to relate the fluid model (3) to the window-based MP-TCP algorithms proposed in the literature. On each route r the source increases its window at the return of each ACK, and let this increment be denoted by $I_r(\mathbf{w}_s)$ where $\mathbf{w}_s$ is the vector of window sizes on different routes of source s. The source decreases the window on route r when it sees a packet loss on route r, and let this decrement by denoted by $D_r(\mathbf{w}_s)$. Let δw_r be the net change to window on route r in each round trip time. Then δw_r is roughly

$$\begin{aligned} \delta w_r &= (I_r(\mathbf{w}_s)(1 - q_r) - D_r(\mathbf{w}_s) q_r) w_r \\ &\approx (I_r(\mathbf{w}_s) - D_r(\mathbf{w}_s) q_r) w_r \end{aligned}$$

when the loss probability q_r is small. On the other hand

$$\delta w_r \approx \dot{w}_r \tau_r = \dot{x}_r \tau_r^2$$

Hence

$$\dot{x}_r = \frac{x_r}{\tau_r} \left(I_r(\mathbf{w}_s) - D_r(\mathbf{w}_s) q_r\right)$$

From (3) we have

$$\begin{cases} k_r(\mathbf{x}_s) &= \frac{2x_r}{\tau_r} D_r(\mathbf{w}_s) \\ \phi_r(\mathbf{x}_s) &= \frac{1}{2}\frac{I_r(\mathbf{w}_s)}{D_r(\mathbf{w}_s)} \end{cases} \quad (5)$$

We now present a fluid model of the existing algorithms in the literature. We will first summarize these algorithms in the form of a pseudo-code and then use (5) to derive parameters $k_r(\mathbf{x}_s)$ and $\phi_r(\mathbf{x}_s)$ of the fluid model (3).

Single-path TCP: TCP-NewReno

Single-path TCP is a special case of MP-TCP algorithm with $|s| = 1$. Hence x_s is a scalar and we identify each source with its route $r = s$. TCP-NewReno adjusts the window as follows:

- Each ACK on route r, $w_r \leftarrow w_r + 1/w_r$.
- Each loss on route r, $w_r \leftarrow w_r/2$.

From (5), this can be modeled by the fluid model (3) with

$$k_r(x_s) = x_r^2, \quad \phi_r(x_s) = \frac{1}{\tau_r^2 x_r^2}$$

We now summarize some existing MP-TCP algorithms, all of which degenerate to TCP NewReno if there is only one route per source.

EWTCP [5]

EWTCP algorithm applies TCP-NewReno like algorithm on each route independently of other routes. It adjusts the window on multiple routes as follows:

- Each ACK on route r, $w_r \leftarrow w_r + a/w_r$.
- Each loss on route r, $w_r \leftarrow w_r/2$.

From (5), this can be modeled by the fluid model (3) with

$$k_r(\mathbf{x}_s) = x_r^2, \quad \phi_r(\mathbf{x}_s) = \frac{a}{\tau_r^2 x_r^2}$$

where $a > 0$ is a constant.

Coupled MPTCP [4, 7]

The Coupled MPTCP algorithm adjusts the window on multiple routes in a coordinated fashion as follows:

- Each ACK on route r, $w_r \leftarrow w_r + \frac{w_r}{(\sum_{k\in s} w_k)^2}$.
- Each loss on route r, $w_r \leftarrow w_r/2$.

From (5), this can be modeled by the fluid model (3) with

$$k_r(\mathbf{x}_s) = x_r^2, \quad \phi_r(\mathbf{x}_s) = \frac{1}{(\sum_{k\in s} x_k\tau_k)^2},$$

Semicoupled MPTCP [13]

The Semi-coupled MPTCP algorithm adjusts the window on multiple routes as follows:

- Each ACK on route r, $w_r \leftarrow w_r + \frac{1}{\sum_{k\in s} w_k}$.
- Each loss on route r, $w_r \leftarrow w_r/2$.

From (5), this can be modeled by the fluid model (3) with

$$k_r(\mathbf{x}_s) = x_r^2, \quad \phi_r(\mathbf{x}_s) = \frac{1}{x_r\tau_r(\sum_{k\in s} x_k\tau_k)}$$

Max MPTCP [13]

The Max MPTCP algorithm adjusts the window on multiple routes as follows:

- Each ACK on route r, $w_r \leftarrow w_r + \min\{\frac{\max\{w_k/\tau_k^2\}}{(\sum w_k/\tau_k)^2}, \frac{1}{w_r}\}$.
- Each loss on route r, $w_r \leftarrow w_r/2$.

From (5), this can be modeled by the fluid model (3) with

$$k_r(\mathbf{x}_s) = x_r^2, \quad \phi_r(\mathbf{x}_s) = \frac{\max\{x_k/\tau_k\}}{x_r\tau_r(\sum_{k\in s} x_k)^2}$$

where we have ignored taking the minimum with the $1/w_r$ term since the performance is mainly captured by $\frac{\max\{w_k/\tau_k^2\}}{(\sum w_k/\tau_k)^2}$.

3. STRUCTURAL PROPERTIES

A point $(\mathbf{x}, \mathbf{p})$ is called an *equilibrium* of (3)–(4) if it satisfies

$$k_r(\mathbf{x}_s)\left(\phi_r(\mathbf{x}_s) - \frac{q_r}{2}\right)_{x_r}^{+} = 0$$
$$\gamma_l\left(y_l - c_l\right)_{p_l}^{+} = 0$$

or equivalently, if

$$\phi_r(\mathbf{x}_s) < \frac{q_r}{2} \Rightarrow \mathbf{x}_r = 0 \text{ and } x_r > 0 \Rightarrow \phi_r(\mathbf{x}_s) = \frac{q_r}{2} \quad (6)$$
$$y_l < c_l \Rightarrow p_l = 0 \text{ and } p_l > 0 \Rightarrow y_l = c_l \quad (7)$$

In this section we identify design criteria that guarantee the existence, uniqueness, and stability of system equilibrium. We characterize algorithm parameters that determine TCP-friendliness and prove an inevitable tradeoff between responsiveness and friendliness. We discuss in the next section the implications of these structural properties on existing algorithms. All proofs are relegated to the Appendix.

3.1 Utility maximization

For single-path TCP (SP-TCP), one can associate a utility function $U_s(x_s) \in \mathbb{R} \to \mathbb{R}$ (x_s is a scalar and $|s| = 1$) with each flow s and interpret (3)–(4) as a distributed algorithm to maximize aggregate user utility [8,10,11], i.e., for SP-TCP, an $(\mathbf{x}, \mathbf{p})$ is an equilibrium if and only if $\mathbf{x}$ is optimal for

$$\text{maximize} \sum_{s\in S} U_s(x_s) \quad \text{s.t. } y_l \le c_l \quad l \in L \quad (8)$$

and $\mathbf{p}$ is optimal for the associated dual problem. Here $y_l \le c_l$ means the aggregate traffic y_l at each link does not exceed its capacity c_l. In fact this holds for a much wider class of SP-TCP algorithms than those specified by (3)–(4) [10]. Furthermore all the main TCP algorithms proposed in the literature have strictly concave utility functions, implying a unique stable equilibrium.

The case of MP-TCP is much more delicate: whether an underlying utility function exists depends on the design choice of Φ_s and not all MP-TCP algorithms have one. Consider the multipath equivalent of (8):

$$\text{maximize} \sum_{s\in S} U_s(\mathbf{x}_s) \quad \text{s.t. } y_l \le c_l \quad l \in L, \quad (9)$$

where $\mathbf{x}_s := (x_r, r \in s)$ is the rate vector of flow s and $U_s(\mathbf{x}_s) \in \mathbb{R}^{|s|} \to \mathbb{R}$ is a concave function. Consider the condition:

C0: The Jacobians of $\Phi_s(\mathbf{x}_s)$ are symmetric for all s, i.e.,

$$\frac{\partial \Phi(\mathbf{x}_s)}{\partial \mathbf{x}_s} = \left[\frac{\partial \Phi(\mathbf{x}_s)}{\partial \mathbf{x}_s}\right]^T$$

Theorem 1. *There exists a twice continuously differentiable $U_s(\mathbf{x}_s)$ such that an equilibrium $(\mathbf{x}, \mathbf{p})$ of (3)–(4) solves (9) and its dual problem if and only if condition C0 holds.*

Condition C0 is satisfied trivially by SP-TCP. For MP-TCP, the models derived in Section 2.2 show that only EWTCP

and Coupled algorithms satisfy C0 and have underlying utility functions. It therefore follows from the theory for SP-TCP that EWTCP has a unique stable equilibrium while Coupled algorithm may have multiple equilibria since its corresponding utility function is not strictly concave.

The other MP-TCP algorithms all have asymmetric Jacobian $\frac{\partial \Phi_s}{\partial \mathbf{x}_s}$. We next study the existence, unique, and stability of these MP-TCP algorithms.

3.2 Existence, uniqueness and stability of equilibrium

Given congestion prices $\mathbf{p} \geq 0$, recall that $\mathbf{q} := H^T\mathbf{p}$. Consider the following conditions:

C1: For each $s \in S$, there exists a nonnegative solution $\mathbf{x}_s := \mathbf{x}_s(\mathbf{p})$ to Eqn. (6) for any $\mathbf{p} \geq 0$. Moreover,

$$\frac{\partial y_l^s(\mathbf{p})}{\partial p_l} \leq 0, \qquad \lim_{p_l \to \infty} y_l^s(\mathbf{p}) = 0,$$

where $y_l^s(\mathbf{p}) := \sum_{r \in s} H_{lr} x_r(\mathbf{p})$, which represents the aggregate traffic at link l from source s.

C2: The Jacobian $\partial \Phi_s(\mathbf{x}_s)/\partial \mathbf{x}_s$ is negative definite and continuous for all $s \in S$.

C3: The routing matrix H has full row rank. For any $r \in R$, $\lim_{x_r \to 0} \phi_r(\mathbf{x}_s) = \infty$ and

$$\sup_{x_{-r} \in \mathbb{R}_+^{|s|-1}} \{\phi_r(\mathbf{x}_s)\} < \infty$$

if $x_r > 0$, where $x_{-r} := \{x_k \mid k \in s \setminus \{r\}\}$.

Condition C1 means that the amount of traffic through a link l from source s does not increase if the congestion price p_l on that link increases. Furthermore, the amount of traffic through that link is 0 if $p_l = \infty$. As mentioned above, the matrix $\partial \Phi_s(\mathbf{x}_s)/\partial \mathbf{x}_s$ in C2 is generally not symmetric. C2 implies that at steady state, if $\mathbf{x}_s, \mathbf{q}_s$ are perturbed by $\delta\mathbf{x}_s, \delta\mathbf{q}_s$ respectively, then $(\delta\mathbf{x}_s)^T \delta\mathbf{q}_s < 0$. When $|s| = 1$, it is equivalent to that the curvature of the utility function is negative, namely $U_s(x_s)$ is strictly concave. Condition C3 means that the rate on route r is zero if and only if it sees infinite price on that route.

The next result says that conditions C1-C3 guarantee that a multipath TCP/AQM (3)–(4) has a unique equilibrium. The proof is given in Appendix A.2.

Theorem 2. *1. Suppose C1 holds. Then (3)–(4) has at least one equilibrium.*

2. Suppose C2 and C3 hold. Then (3)–(4) has at most one equilibrium

Thus (3)–(4) has a unique equilibrium $(\mathbf{x}^*, \mathbf{p}^*)$ under C1–C3.

Conditions C1-C3 not only guarantee the existence and uniqueness of the equilibrium, they also ensure that the equilibrium is globally asymptotically stable, when the gain $k_r(\mathbf{x}_s)$ is a constant, i.e., $k_r(\mathbf{x}_s) \equiv k_r$ for all $r \in R$.

Theorem 3. *Suppose C1-C3 hold and $k_r(\mathbf{x}_s) \equiv k_r$ for all $r \in R$. Starting from any initial point $\mathbf{x}(0) \in \mathbb{R}_+^{|R|}$ and $\mathbf{p}(0) \in \mathbb{R}_+^{|L|}$, the trajectory $(\mathbf{x}(t), \mathbf{p}(t))$ generated by the MP-TCP algorithm (3)–(4) converges to the unique equilibrium $(\mathbf{x}^*, \mathbf{p}^*)$ as $t \to \infty$.*

Since we can treat the gain $k_r(\mathbf{x}_s)$ as a constant if we linearize the system (3)–(4) around the equilibrium $(\mathbf{x}^*, \mathbf{p}^*)$. Theorem 3 can also serve as a proof of $(\mathbf{x}^*, \mathbf{p}^*)$ as a local asymptotically stabile equilibrium.

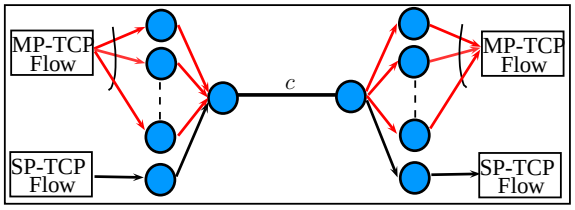

Figure 1: Test network for the definition of TCP friendliness. The link in the middle is the only bottleneck link with capacity c.

3.3 TCP Friendliness

Informally, an MP-TCP flow is said to be 'TCP friendly' if it does not dominate the available bandwidth when it shares the same network with a SP-TCP flow [3]. To define this precisely we use the test network shown in Fig. 1 where there are a fixed number of paths all traversing a single bottleneck link with capacity c. All other links have capacities strictly higher than c. The links have fixed but possibly different delays. The test network is shared by two flows, one SP-TCP and the other MP-TCP under test as shown in the figure. To compare the friendliness of two MP-TCP algorithms $M1$ and $M2$, suppose that when $M1$ shares the test network with the SP-TCP it achieves a throughput of $||\mathbf{x}_1||_1$ in equilibrium aggregated over the available paths. Suppose $M2$ achieves a throughput of $||\mathbf{x}_2||_1$ in equilibrium when it shares the test network with the SP-TCP. Then we say that $M1$ is *more friendly* than $M2$ if $||\mathbf{x}_1||_1 \leq ||\mathbf{x}_2||_1$, i.e., if $M1$ receives no more bandwidth than $M2$ does when they *separately* share the test network in Figure 1 with the SP-TCP flow.

Let $D_s := \left[\frac{\partial \Phi_s(\mathbf{x}_s)}{\partial \mathbf{x}_s}\right]^{-1}$, whose existence is guaranteed by C2. Consider the following design criterion for Φ_s:

C4 : For all routes $r \in s$, $\sum_{j \in s} [D_s]_{jr} \leq 0$.

To interpret C4, note that the equilibrium condition imposes $\Phi_s(\mathbf{x}_s) = \frac{1}{2}\mathbf{q}_s$. The implicit function theorem then implies $\mathbf{1}^T \frac{\partial \mathbf{x}_s}{\partial q_r} = \sum_{j \in s} D_{jr}$ for all $r \in s$. Hence C4 says that the aggregate throughput $\mathbf{1}^T\mathbf{x}_s$ at equilibrium of a MP-TCP flow is a nonincreasing function of the price q_r on all routes $r \in s$.

The next result says that an MP-TCP algorithm is more TCP friendly if it has a smaller $\Phi_s(\mathbf{x}_s)$.

Theorem 4. *Consider two MP-TCP algorithms $M1$ with $\hat{\Phi}_s(\mathbf{x}_s)$ and $M2$ with $\tilde{\Phi}_s(\mathbf{x}_s)$. Suppose both satisfy C1–C4. Then $M1$ is more friendly than $M2$ if $\hat{\Phi}_s(\mathbf{x}_s) \leq \tilde{\Phi}_s(\mathbf{x}_s)$.*

3.4 Responsiveness

The linearized system of (3)–(4) at the equilibrium $(\mathbf{x}^*, \mathbf{p}^*)$ is defined by the Jacobian

$$J^* := \begin{bmatrix} \Lambda_k \frac{\partial \Phi}{\partial \mathbf{x}} & -\frac{1}{2}\Lambda_k H^T \\ \Lambda_\gamma H & 0 \end{bmatrix}$$

where $\Lambda_k = diag\{k_r(\mathbf{x}_s^*), r \in s\}$, $\Lambda_\gamma = diag\{\gamma_l, l \in L\}$ and $\frac{\partial \Phi}{\partial \mathbf{x}}$ is evaluated at $\mathbf{x}^*$. The stability and responsiveness (how fast does the system converges to the equilibrium locally) of the system (3)–(4) is determined by the real part of the eigenvalues of J^*. Specifically the linearized system is asymptotically stable if the real parts of all the eigenvalues of J^* are negative; moreover the more negative the real parts are the faster the linearized system converges to the equilibrium.

We are therefore interested in an upper bound on the largest (least negative) real part of the eigenvalue of J^* in

terms of both K_s and $\frac{\partial \Phi_s}{\partial \mathbf{x}_s}$. Without loss of generality, assume all the links in L are active with $p_l^* > 0$. Otherwise, we can remove all the links with price $p_l^* = 0$ from our model.

Let $\lambda(J^*)$ be the set of eigenvalues of J and define the largest real part of eigenvalue:

$$\lambda_m(J^*) := \max\{\mathbf{Re}(\lambda) \mid \lambda \in \lambda(J^*)\}$$

Let $\left[\frac{\partial \Phi}{\partial \mathbf{x}}\right]^+ := \frac{1}{2}\left(\frac{\partial \Phi}{\partial \mathbf{x}} + \left[\frac{\partial \Phi}{\partial \mathbf{x}}\right]^T\right)$ and $S = \{(\mathbf{z}, \mathbf{p}) \mid \|\mathbf{z}\|_2 = \|\mathbf{p}\|_2 = 1, \mathbf{z} \in \mathbb{C}^{|R|}, \mathbf{p} \in \mathbb{C}^{|L|}\}$.

Lemma 1. *Suppose C2 holds. Then*

$$\lambda_m(J^*) \le \overline{\lambda}_J := \max_{(\mathbf{z},\mathbf{p})\in S} \left\{ \frac{2\mathbf{z}^H \left[\frac{\partial \Phi}{\partial \mathbf{x}}\right]^+ \mathbf{z}}{2\mathbf{z}^H \Lambda_k^{-1} \mathbf{z} + \mathbf{p}^H \Lambda_\gamma^{-1} \mathbf{p}} \right\}$$

To understand the implication of the lemma, consider two MP-TCP algorithms $(\hat{K}_s, \hat{\Phi}_s)$ and $(\tilde{K}_s, \tilde{\Phi}_s)$. If $\hat{\Lambda}_k \ge \tilde{\Lambda}_k$ and $\frac{\partial \hat{\Phi}}{\partial \mathbf{x}} \preceq \frac{\partial \tilde{\Phi}}{\partial \mathbf{x}}$ then

$$\mathbf{z}^H \hat{\Lambda}_k^{-1} \mathbf{z} \le \mathbf{z}^H \tilde{\Lambda}_k^{-1} \mathbf{z} \quad (10)$$

$$\mathbf{z}^H \left[\frac{\partial \hat{\Phi}}{\partial \mathbf{x}}\right]^+ \mathbf{z} \le \mathbf{z}^H \left[\frac{\partial \tilde{\Phi}}{\partial \mathbf{x}}\right]^+ \mathbf{z} \quad (11)$$

for any $\|\mathbf{z}\|_2 = 1$ and thus $\overline{\lambda}_{\hat{J}} \le \overline{\lambda}_{\tilde{J}}$. Note that Λ_k is a diagonal matrix consisting of $k_r(\mathbf{x}_s^*)$ for all $r \in R$ and $\partial\Phi/\partial\mathbf{x}$ is a block diagonal matrix consisting of $\partial\Phi_s/\partial\mathbf{x}_s$ for all $s \in S$. Then Eqn. (10)-(11) hold if $\hat{K}_s \ge \tilde{K}_s$ and $\frac{\partial \hat{\Phi}_s}{\partial \mathbf{x}_s} \preceq \frac{\partial \tilde{\Phi}_s}{\partial \mathbf{x}_s}$. Hence, informally Lemma 1 says that an MP-TCP algorithm with a larger $K_s(\mathbf{x}_s)$ or more negative definite $\frac{\partial \Phi}{\partial \mathbf{x}}$ is more responsive, in the sense that the real parts of the eigenvalues of the Jacobina J^* have a smaller upper bound.

Then the next result suggests an inevitable tradeoff between responsiveness and un-friendliness. Consider

C5 : For all routes $r \in s$, $\phi_r(\mathbf{x}_s) \le (x_r\tau_r)^{-2}$.

For SP-TCP, $\phi_r(x_s) = (x_r\tau_r)^{-2}$ as shown in Section 2.2. C5 simply means that on any route $r \in s$, a MP-TCP flow is no more aggressive than a SP-TCP flow.

Theorem 5. *Consider two MP-TCP algorithms with $\hat{\Phi}_s(\mathbf{x}_s)$ and $\tilde{\Phi}_s(\mathbf{x}_s)$. Suppose both satisfy C1-C3 and C5. Then*

$$\frac{\partial \hat{\Phi}_s(\mathbf{x}_s)}{\partial \mathbf{x}_s} \preceq \frac{\partial \tilde{\Phi}_s(\mathbf{x}_s)}{\partial \mathbf{x}_s} \Rightarrow \hat{\Phi}_s(\mathbf{x}_s) \ge \tilde{\Phi}_s(\mathbf{x}_s)$$

In light of Lemma 1 and Theorem 4, Theorem 5 says that a more responsive MP-TCP design is inevitably less friendly.

4. IMPLICATION AND A NEW ALGORITHM

TCP friendliness, responsiveness and window fluctuation are crucial performance metrics of TCP algorithms. We have already studied TCP friendliness and responsiveness in section 3 under the unified fluid model (3)-(4). In this section, we will study the relation between window fluctuation, which cannot be captured in the fluid model, and the design parameters. Then we will discuss the implications of the structural properties proved in Section 3 on the behavior of existing MP-TCP algorithms. The discussions motivate a new design that generalizes and extends the existing MP-TCP algorithm. We present our design rationale and further illustrate its performance through simulations in Section 5.

4.1 Window Fluctuation

For loss based TCP algorithm, the congestion window always fluctuates due to the binary congestion signal and can not be avoided. It can not be captured in the fluid model (3)-(4). Therefore, the equilibrium specified by the fluid dynamics (3)-(4) represents the average realtime throughput. For TCP-NewReno, the throughput is halved in the next RTT if one packet loss is observed - we define throughput as the window size over the round trip time. Hence, the fraction of throughput reduction is 1/2. *Window fluctuation* is a measure of how the throughput of the source fluctuates due to the binary congestion signal. In this paper, we use $\|K_s(\mathbf{x}_s)\|_1$ as a measure of the window fluctuation and smaller $\|K_s(\mathbf{x}_s)\|_1$ means smaller *window fluctuation*. A heuristic is given below that motivates this measure.

Suppose the loss probability q_r are the same for all $r \in s$. When there is one packet loss on $r \in s$, its window will decrease by $\frac{k_r(\mathbf{x}_s)\tau_r}{2x_r}$ based on Eqn. (5), which means throughput will decrease by $\frac{k_r(\mathbf{x}_s)}{2x_r}$, for each packet loss on route r. Note that the loss probability is approximately $x_r q_r$ on route r. Then each time a packet loss is detected by source s, the probability that the loss happens at route j is approximately

$$\frac{x_j q_j}{\sum_{r\in s} x_r q_r} = \frac{x_j}{\|\mathbf{x}_s\|_1},$$

provided q_r is small and the same for all routes $r \in s$. Hence, the average throughput reduction over the aggregate throughput $\|\mathbf{x}_s\|_1$ is

$$\frac{1}{\|\mathbf{x}_s\|_1} \sum_{r\in s} \frac{x_r}{\|\mathbf{x}_s\|_1} \frac{k_r(\mathbf{x}_s)}{2x_r} = \frac{\|K_s(\mathbf{x}_s)\|_1}{2\|\mathbf{x}_s\|_1^2}, \quad (12)$$

which means smaller $\|K_s(\mathbf{x}_s)\|_1$ leads to smaller throughput reduction per packet loss.

For the existing algorithms, all of them have $k_r(\mathbf{x}_s) = x_r^2$ thus their fluctuation level are the same under our metrics. Substitute $k_r(\mathbf{x}_s) = x_r^2$ into Eqn. (12), we find that it is always smaller than 1/2. Since the the average throughput reduction is 1/2 for TCP-NewReno by Eqn. (12), existing MP-TCP algorithms always have better window fluctuation performance than TCP NewReno. Indeed, enabling MP-TCP always improves the window fluctuation performance provided $k_r(\mathbf{x}_s) \le x_r\|\mathbf{x}_s\|_1$. Recall that larger $K_s(\mathbf{x}_s)$ means better responsiveness performance as shown in 3.4. It means there is tradeoff between responsiveness and window fluctuation, which can be improved by using smaller $K_s(\mathbf{x}_s)$.

Table 1: MP-TCP Algorithms

Algorithm	C0	C1	C2-C3	C4	C5
EWTCP	Yes	Yes	Yes	Yes	Yes
Coupled	Yes	Yes	No	Yes	Yes
Semicoupled	No	Yes	Yes	Yes	Yes
Max	No	Yes	Yes	Yes	Yes
Our algorithm	No	Yes	Yes	Yes	Yes

4.2 Implications on existing algorithms

In this subsection, we will study the equilibrium, TCP friendliness and responsiveness property of existing algorithms introduced in section 2.2. They can serve as a verification of our analysis in section 3. We have briefly summarized existing algorithms together with our algorithm, which will be proposed in the next subsection, in Table 1.

First, we will study whether there exists a unique stable equilibrium to the existing algorithms. We find that only EWTCP and Coupled algorithm satisfies C0, thus their equilibrium property can be studied in the context of utility maximization as shown in Theorem 1. However, Semicoupled and Max algorithm does not satisfy C0 and thus we indeed need to rely on Theorem 2 and 3 to study their

Table 2: Design Space

Performance	Deterministic parameter
TCP Friendliness	$\phi_r(\mathbf{x}_s) \uparrow$
Responsiveness	$k_r(\mathbf{x}_s) \uparrow$, $-\partial\Phi_s/\partial\mathbf{x}_s \uparrow$
Window Fluctuation	$\|K_s(\mathbf{x}_s)\|_1 \downarrow$

equilibrium performance by showing whether C1-C3 are satisfied. For EWTCP, it is easy to show they satisfy C1-C3. For the other existing algorithms, they can be modeled using Eqn. (13) with specified parameters.

$$\begin{cases} k_r(\mathbf{x}_s) &= x_r(x_r + \eta(\|\mathbf{x}_s\|_\infty - x_r)). \quad \eta \geq 0 \\ \phi_r(\mathbf{x}_s) &= \frac{x_r + \beta(\|\mathbf{x}_s\|_n - x_r)}{\tau_r^2 x_r \|\mathbf{x}_s\|_1^2}, \quad n \in \mathbb{N}_+, \beta \geq 0 \end{cases} \tag{13}$$

Max algorithm ($\beta = 1, n = \infty, \eta = 0$), Semicoupled algorithm ($\beta = 1, n = 1, \eta = 0$)[2] and Coupled algorithm ($\beta = 0, \eta = 0$) are special cases corresponding to different (β, n, η). The next theorem shows whether they satisfy C1-C3 by studying the general (13).

Theorem 6. *For any $n \in \mathbb{N}_+$, the $\phi_r(\mathbf{x}_s)$ proposed in Eqn. (13) satisfies C1 if $0 \leq \beta \leq 1$. Furthermore, it satisfies C2-C3 if $0 < \beta \leq 1$, $|s| \leq 8$ and τ_r is the same across all $r \in s$.*

Therefore, when the round trip time of each route is the same, there exits a unique stable equilibrium to Max and Semicoupled algorithm if they enable less than 8 routes. For Coupled algorithm, it does not satisfy C2-C3 and is found to have multiple equilibria in [7].

For a negative definite matrix A, it is still negative definite after some small perturbations of each entry. Therefore, Theorem 6 holds provided the RTT of each subpath does not differ much. Indeed, the round trip time of each subpath is close in reality since it is mainly determined by the distance between two hosts. In the proof of Theorem 6, the Jacobian of $\Phi_s(\mathbf{x}_s)$ becomes less negative definite when the number of routes $|s|$ increases and will finally be indefinite when $|s| > 9$. Thus smaller $|s|$ offers larger freedom of RTT heterogeneity across each route to maintain its negative definite property. However, it is a less contingent condition since each source is typically enabled 2 or 3 routes in reality.

Second, we will study the friendliness performance of existing MP-TCP algorithms. In Theorem 4, we show that algorithm with larger $\phi_r(\mathbf{x}_s)$ is less TCP friendly. For the existing MP-TCP algorithm, all of them satisfy C4 and their throughput for the network in Fig. 1 is ordered as follows:

$$\text{EWTCP}(a \geq 1)^3 \geq \text{Semicoupled} \geq \text{Max} \geq \text{Coupled}$$

Since the more bandwidth the MP-TCP source occupies, the more aggressive the algorithm is. It means EWTCP is the most aggressive while Coupled algorithm is the most fair one. On the other hand, the $\phi_r(\mathbf{x}_s)$ corresponding to each algorithm in section 2.2 is ordered as follows:

$$\phi_r^{ewtcp}(\mathbf{x}_s) \geq \phi_r^{semicoupled}(\mathbf{x}_s) \geq \phi_r^{max}(\mathbf{x}_s) \geq \phi_r^{coupled}(\mathbf{x}_s)$$

for all $\mathbf{x}_s \geq 0$ if each route has the same round trip time. Thus it verifies that algorithm with larger $\phi_r(\mathbf{x}_s)$ tends to be less TCP friendly as shown in Theorem 4.

Third, we will study the responsiveness performance of existing MP-TCP algorithms. Lemma 1 informally say that an MP-TCP algorithm with a larger $K_s(\mathbf{x}_s)$ or more negative definite $\frac{\partial\Phi_s}{\partial\mathbf{x}_s}$ is more responsive. For prior algorithms, they have the same gain function $k_r(\mathbf{x}_s) = x_r^2$ and

$$(\frac{\partial\Phi_s}{\partial\mathbf{x}_s})^{ewtcp} \preceq (\frac{\partial\Phi_s}{\partial\mathbf{x}_s})^{semicoupled} \preceq (\frac{\partial\Phi_s}{\partial\mathbf{x}_s})^{max} \preceq (\frac{\partial\Phi_s}{\partial\mathbf{x}_s})^{coupled}.$$

Out simulations in section 5 show that the responsiveness performance is the same as the above order. It means that EWTCP is the best algorithm in terms of responsiveness while Coupled algorithm, whose $\frac{\partial\Phi_s}{\partial\mathbf{x}_s}$ is merely negative semi-definite, is the worst.

[2]The constant in front of the variable $\mathbf{x}_s$ in the algorithms are different. But the analysis for showing whether they satisfy C1-C3 can be carried on in a similar manner.

[3]When $a < 1$, the MP-TCP source will obtain even less throughput than competing single-path TCP source sometimes.

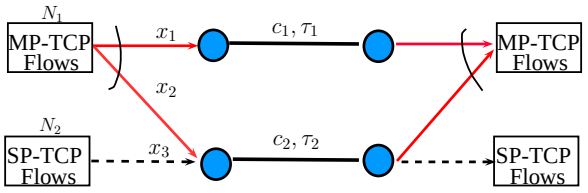

Figure 2: Network with N_1 MP-TCP flows and N_2 single-path TCP flows sharing 2 links of capacity c_1, c_2 and delay τ_1, τ_2. MP-TCP flows maintain two routes with rate x_1, x_2. Single-path TCP flows maintain one route with rate x_3.

4.3 A new MP-TCP design

We have discussed the relation between different performances and the corresponding parameters, which are summarized in Table 2. As discussed above the design of MP-TCP algorithms involves inevitable tradeoffs. For instance a more negative definite $\partial\Phi_s/\partial\mathbf{x}_s$ enhances responsiveness but is less friendly to SP-TCP; a higher gain $k_r(\mathbf{x}_s)$ usually improves responsiveness but is more oscillatory. Therefore it is impossible to have an algorithm that is superior for all performances.

As mentioned in section 4.1, window fluctuation is improved compared to using SP-TCP if $k_r(\mathbf{x}_s) \leq x_r\|\mathbf{x}_s\|_1$ and most current algorithms use $k_r(\mathbf{x}_s) = x_r^2$. Therefore we can sacrifice the window fluctuation performance a bit while boost up responsiveness by using a relative large $k_r(\mathbf{x}_s)$. Then it leaves us more space to improve the fairness by using small $\phi_r(\mathbf{x}_s)$. Furthermore, responsiveness is mainly affected by subpaths with small throughput while stability is mainly affected by subpaths with large throughput. We will keep $k_r(\mathbf{x}_s) = x_r^2$ nearly unchanged, which is used by default in prior algorithms, on route with large throughput but increase increase $k_r(\mathbf{x}_s)$ on route with small throughput.

Now we are left with developing a parameterized candidate, which can be tuned to reach different region in the design space. We will use the generalized algorithm of Eqn. (13) and specify the parameters (β, n, η). As discussed above, we need to choose big $k_r(\mathbf{x}_s)$ and small $\phi_r(\mathbf{x}_s)$. Note that the algorithm is computationally efficient when $n = 1$ or ∞ since there is no exponentiation. Considering these choices and the experiments in ns2, we pick $(\beta, n, \eta) = (0.2, \infty, 0.5)$. The corresponding algorithm satisfies C1-C3 based on Theorem 6 and they also satisfy C4-C5, whose proof is skipped due to space limitation. Then we can convert it into window based algorithm according to Eqn. (5), which corresponds to our algorithm at the end of section 1. Next, we will test the performance of our algorithm using ns2 simulations to confirm our analysis.

5. SIMULATION

In this section we briefly summarize our ns2 simulation results and compare the performance of our algorithm with prior algorithms. It also serves to confirm our theoretical analysis. In the simulations we set the slow start phase on each route the same as of TCP-NewReno; however, the min-

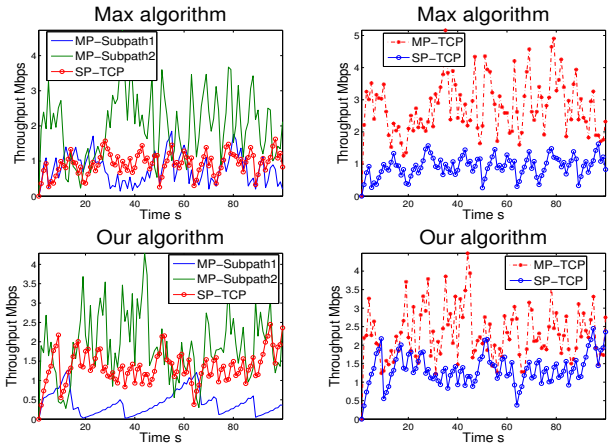

Figure 3: The throughput of MP-TCP and single-path TCP users with network topology in Fig. 2 and different RTT. The figures on the left show the throughput of each route and figures on the right show the total throughput of each source.

Table 3: Throughput of MPTCP and single-path TCP users.

	ewtcp	semi.	max	ours	coupled
MP-TCP Mbps	2.98	2.64	2.58	2.25	2.22
SP-TCP Mbps	1.01	1.32	1.35	1.61	1.67

imum *ssthresh* is set to 1 instead of 2 when there are more than 1 routes available. When the congestion avoidance phase starts, the congestion window size *cwnd* is adapted as stated in the algorithms for each subpath. We assume the advertised window *awnd* is set to be infinity.

Our simulations are divided into three parts. First, we compare the friendliness performance of our algorithm and prior algorithms. When the round trip time of each subpath is the same, we show that our algorithm is close to that of Coupled algorithm, which is the best in terms of TCP friendliness, and outperform the other prior algorithms. When the round trip time of each subpath is different, our algorithm also works well while prior algorithms, e.g. Max algorithm, is more aggressive. Second, we compare the responsiveness performance of each algorithm when there are users come and go. We show that Coupled algorithm does not work well in the dynamic environment, while our algorithm is as responsive as the other algorithms and, unlike the other algorithms, is more friendly to SP-TCP flows. Finally, we show that our algorithm achieves better window fluctuation performance than single-path TCP. Consider the experiment we have done about TCP friendliness, responsiveness and window fluctuation performance, we claim that our algorithm gives a better overall balance performance.

5.1 TCP Friendliness Performance

In this subsection, we will study the TCP friendliness performance of each algorithm using the network topology in Fig. 2. We assume all the flows are long lived and focus on the steady state throughput. For the network shown in Fig. 2, let $\tau_1 = \tau_2 = 20$ms, $c_1 = c_2 = 10$Mbps and $N_1 = N_2 = 5$. The average aggregate throughput of MP-TCP and single-path TCP users are shown in Table 3.

According to Table 3, EWTCP, Semicoupled and Max algorithm are very aggressive and severely harm the single-path TCP users, our algorithm is close to Coupled algorithm and is good in terms of TCP friendliness. Indeed, both MP-TCP and SP-TCP users should obtain 2Mbps throughput in the ideal friendly case, which means there is no traffic on the second route for all the MP-TCP users. However, since

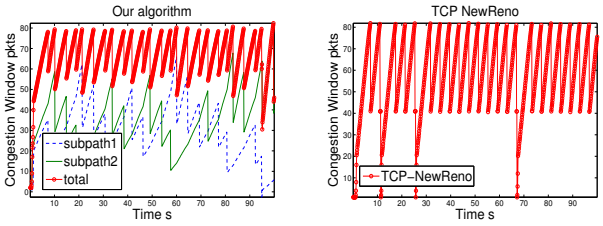

Figure 4: The window fluctuation. The figure on the left shows our algorithm and the right one is TCP-NewReno.

Table 4: Convergence Time for MP-TCP and throughput of single-path TCP

	ewtcp	semi.	max	ours	coupled
ConvergeTime s	1	2.5	4.5	4.5	54
Througput Mbps	1.02	1.17	1.30	1.57	1.72

the minimal window size is 1 and even Coupled algorithms, which is the most fair one, also exhibit some extent of aggressiveness.

We have already shown how MP-TCP users affect the SP-TCP users for different algorithms when the round trip time of the subpath are similar. We now show that prior algorithms will be less friendly to SP-TCP when the round trip time of each subpath is different. Due to space limitation, we only show the results obtained under our algorithm and Max algorithm. Using network topology in Fig. 2 with $c_1 = c_2 = 10$Mbps, $N_1 = N_2 = 5$. But the propagation delay of the two links are different with $\tau_1 = 10$ms and $\tau_2 = 50$ms. Then the RTT for the first and second routes are roughly 20ms and 100ms. The simulation results are shown in Fig. 3. We obverse that our algorithm is quite moderate in comparison with Max algorithm.

5.2 Responsiveness Performance

In this subsection, we compare the responsiveness performance of each algorithm. Since most of the traffic in the current Internet is mice traffic, a good algorithm needs to react fast enough in this dynamic environment. We still use the network topology shown in Fig. 2 with $c_1 = c_2 = 2$Mbps, $\tau_1 = \tau_2 = 20$ms and $N_1 = N_2 = 1$. We assume that the MP-TCP flow is long lived, while the SP-TCP flow starts at 40s and end at 80s. We measure the average throughput of that single-path TCP flow from 40-80s, which reflects the aggressiveness of MP-TCP. We also measure the time for the route on the second link to recover[4] for MP-TCP users. It reflects the responsiveness performance of MP-TCP and a responsive algorithm should have small recover time. The trajectories of the throughput are shown in Fig. 5 for prior and our algorithms. The average throughput of that SP-TCP flow and the time for the MP-TCP user to recover when the single-path TCP flow leaves the network are shown in Table 4.

According to Table 4, EWTCP is the best in terms of responsiveness, and our algorithm is as responsive as the Max algorithm. It takes about 54s for the Coupled algorithm to recover, which is a serious drawback. As to the throughput, which determines friendliness property, our algorithm is still close to that of Coupled algorithm.

5.3 Window Fluctuation Performance

Finally, we consider the window fluctuation performance. We assume a toy model where there is only one link in the network. We consider two experiments. In the first

[4]Defined as the first time the throughput on the second route is within 90% of the average throughput after the single-path user leave.

experiment, our MP-TCP algorithm initiates two subpaths through that link, we monitor the window size of each subpath and their aggregate window size. In the second experiment, let TCP-NewReno algorithm traverse the same link and we monitor the window size. The results are shown in Fig. 4, which demonstrates that our algorithm is better than TCP-NewReno in terms of window fluctuation performance. And prior MP-TCP algorithms have similar property, that their window fluctuation performance is better than TCP NewReno, which confirms our analysis in section 4.1.

6. CONCLUSION

In this paper, we investigate the emerging problems in designing MP-TCP congestion control algorithms. We show the existence, uniqueness and stability properties of the equilibria under a unified model of MP-TCP algorithms using a new approach. Furthermore, we characterize the design space and study the tradeoffs among various performance metrics. Finally, we design a new MP-TCP algorithm and compare its performance with existing algorithms through ns2 simulation.

7. REFERENCES

[1] S. Boyd and L. Vandenberghe. Convex optimization. *Cambridge university press*, 2004.

[2] C. Cetinkaya and E. W. Knightly. Opportunistic traffic scheduling over multiple network paths. In *INFOCOM 2004.*, volume 3, pages 1928–1937. IEEE, 2004.

[3] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure. Tcp extensions for multipath operation with multiple addresses. *IETF MPTCP proposal*, 2009.

[4] H. Han, S. Shakkottai, C. Hollot, R. Srikant, and D. Towsley. Overlay tcp for multi-path routing and congestion control. In *IMA Workshop on Measurements and Modeling of the Internet*, 2004.

[5] M. Honda, Y. Nishida, L. Eggert, P. Sarolahti, and H. Tokuda. Multipath congestion control for shared bottleneck. In *Proc. PFLDNeT workshop*, 2009.

[6] J. R. Iyengar, P. D. Amer, and R. Stewart. Concurrent multipath transfer using sctp multihoming over independent end-to-end paths. *Networking, IEEE/ACM Transactions on*, 14(5):951–964, 2006.

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

[8] F. P. Kelly, A. K. Maulloo, and D. K. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, 49(3):237–252, 1998.

[9] W. S. Lohmiller and J.-J. E. Slotine. *Contraction analysis of nonlinear systems*. PhD thesis, Massachusetts Institute of Technology, Dept. of Mechanical Engineering, 1999.

[10] S. H. Low. A duality model of tcp and queue management algorithms. *Networking, IEEE/ACM Transactions on*, 11(4):525–536, 2003.

[11] S. H. Low and D. E. Lapsley. Optimization flow control-I: basic algorithm and convergence. *IEEE/ACM Transactions on Networking (TON)*, 7(6):861–874, 1999.

[12] A. Tang, J. Wang, and S. H. Low. Is fair allocation always inefficient. In *INFOCOM 2004.*, volume 1. IEEE, 2004.

[13] D. Wischik, C. Raiciu, A. Greenhalgh, and M. Handley. Design, implementation and evaluation of congestion control for multipath tcp. In *Proceedings of the 8th USENIX conference on Networked systems design and implementation*, pages 8–8. USENIX Association, 2011.

[14] 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. In *Proceedings of the annual conference on USENIX Annual Technical Conference*, pages 8–8. USENIX Association, 2004.

8. ACKNOWLEDGMENTS

This work was supported by ARO MURI through grant W911NF-08-1-0233, NSF NetSE through grant CNS 0911041, Bell Labs, Lucent-Alcatel and Wilfred Kwan.

APPENDIX

A. PROOF

Some math notations are summarized. $\mathbf{1}$ is an all one vector whose dimension is clear in the context. I is the identity matrix.

A.1 Proof of Theorem 1

The Lagrangian of (9) is:

$$\begin{aligned} L(\mathbf{x},\mathbf{p}) &= \sum_{s\in S} U_s(\mathbf{x}_s) - \sum_{l\in L} p_l(y_l - c_l) \\ &= \sum_{s\in S} U_s(\mathbf{x}_s) - \sum_{l\in L} p_l(\sum_{r\in R} H_{lr}x_r - c_l) \\ &= \sum_{s\in S}\left(U_s(\mathbf{x}_s) - \sum_{r\in s} x_r q_r\right) + \sum_{l\in L} p_l c_l \end{aligned}$$

where $\mathbf{p} \geq \mathbf{0}$ are the dual variables and $q_r := \sum_{r\in R} H_{lr}p_l$. Then the dual problem is

$$D(\mathbf{p}) = \sum_{s\in S} \max_{\mathbf{x}_s \geq 0}\{B_s(\mathbf{x}_s,\mathbf{p})\} + \sum_{l\in L} p_l c_l \quad \mathbf{p}\geq \mathbf{0}$$

where $B_s(\mathbf{x}_s,\mathbf{p}) = U_s(\mathbf{x}_s) - \sum_{r\in s} x_r q_r$. The KKT condition implies that, at optimality, we have

$$\frac{\partial U_s(\mathbf{x}_s)}{\partial x_r} < q_r \Rightarrow x_r = 0 \text{ and } x_r > 0 \Rightarrow \frac{\partial U_s(\mathbf{x}_s)}{\partial x_r} = q_r \quad (14)$$

$$y_l < c_l \Rightarrow p_l = 0 \text{ and } p_l > 0 \Rightarrow y_l = c_l \quad (15)$$

Comparing with (6)–(7) we conclude that, if a MP-TCP algorithm defined by (3)–(4) has an underlying utility function U_s, then we must have

$$\frac{\partial U_s(\mathbf{x}_s)}{\partial x_r} = 2\phi_r(\mathbf{x}_s) \quad r\in s,\ x_r > 0 \quad (16)$$

Given $\phi_r(\mathbf{x}_s)$, (16) has a continuously differentiable solutions $U_s(\mathbf{x}_s)$ if and only if the Jacobian of $\Phi_s(\mathbf{x}_s)$ is symmetric, i.e., if and only if

$$\frac{\partial \Phi(\mathbf{x}_s)}{\partial \mathbf{x}_s} = \left(\frac{\partial \Phi(\mathbf{x}_s)}{\partial \mathbf{x}_s}\right)^T$$

A.2 Proof of Theorem 2

Proof of Part I:

For any link l, let $p_{-l} = \{p_1,\ldots,p_{l-1},p_{l+1},\ldots,p_{|L|}\}$, whose component composes of all the elements in $\mathbf{p}$ except p_l. In

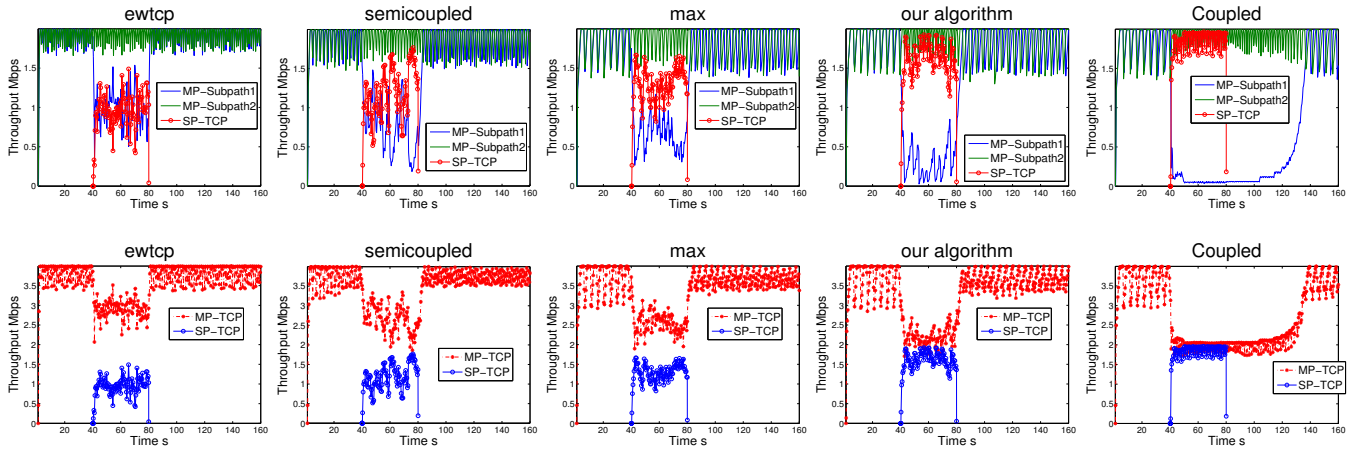

Figure 5: The dynamic behavior of MP-TCP algorithm. The topology is in Fig. 2 with long lived MP-TCP flows and short single-path TCP flows. The upper figures are the rate trajectory for each path and the lower ones are the trajectory of the total throughput for users with MP-TCP and single-path TCP algorithms.

other words, $\mathbf{p} = (p_l, p_{-l})$ for any $l \in L$. For $l \in L$, let

$$g_l(\mathbf{p}) := c_l - \sum_{r:l\in r} x_r = c_l - \sum_{s:r\in s, l\in r} y_l^s(p_l, p_{-l})$$

and $h_l(\mathbf{p}) = -g_l^2(\mathbf{p})$. According to *C1*, we have the following two facts, which will be used in the proof.

- $g_l(\mathbf{p})$ is a nondecreasing function of p_l on $\mathbb{R}_+$ since $y_l^s(\mathbf{p})$ is a nonincreasing function of p_l.
- $\lim_{p_l\to\infty} g_l(p_l, p_{-l}) = c_l$ since $\lim_{p_l\to\infty} y_l^s(\mathbf{p}) = 0$.

Next, we will show that $h_l(\mathbf{p})$ is a quasi-concave function of p_l. In other words, for any fixed p_{-l}, the set $S_a := \{p_l \mid h_l(\mathbf{p}) \geq a\}$ is a convex set. If $g_l(0, p_{-l}) \geq 0$, then

$$g_l(p_l, p_{-l}) \geq g_l(0, p_{-l}) \geq 0 \quad \forall p_l \geq 0,$$

which means $h_l(p_l, p_{-l})$ is a nonincreasing function of p_l, hence is a quasi-concave function of p_l and

$$\arg\max_{p_l} h_l(p_l, p_{-l}) = 0. \tag{17}$$

On the other hand, if $g_l(0, p_{-l}) < 0$, then there exists a $p_l^* > 0$ such that $g_l(p_l^*, p_{-l}) = 0$ since $g_l(\cdot)$ is continuous and $\lim_{p_l\to\infty} g_l(p_l, p_{-l}) = c_l > 0$. Note that $g_l(\mathbf{p})$ is a nondecreasing function of p_l, then $h_l(p_l, p_{-l})$ is nondecreasing for $p_l \in [0, p_l^*]$ and nonincreasing for $p_l \in [p_l^*, \infty)$. Hence, $h_l(p_l, p_{-l})$ is also a quasi-concave function of p_l in this case and

$$\max_{p_l} h_l(p_l, p_{-l}) = 0. \tag{18}$$

By Nash theorem, if $h_l(p_l, p_{-l})$ is a quasi-concave function of p_l for all $l \in L$ and $\mathbf{p}$ is in a bounded set, then there exists a $\mathbf{p}^\star \in \mathbb{R}_+^{|L|}$ such that

$$p_l^\star = \arg\max_{p_l\in\mathbb{R}_+} h_l(p_l, p_{-l}^\star).$$

According to Eqn. (17) and (18), for any $l \in L$, either $p_l^* > 0$ or $g_l^*(\mathbf{p}^*) > 0$ but not both holds at any time. Therefore $\mathbf{p}^*$ satisfies Eqn. (7). Since $\mathbf{q} = R^T\mathbf{p}$, there exists a $\mathbf{x}^*$ to Eqn. (6). And we arrive at our conclusion that there exists at least one solution $(\mathbf{x}, \mathbf{p})$ which satisfies Eqn.(6) and (7).

Proof of Part II:

Lemma 2. *Assume function $F(\mathbf{x}) \in \mathbb{R}^n \to \mathbb{R}^n$ and $\partial F/\partial \mathbf{x}$ is negative definite and continuous in $\mathbb{R}^n$. Then for any $\mathbf{x}_1 \neq \mathbf{x}_2 \in \mathbb{R}^n$,*

$$(\mathbf{x}_1 - \mathbf{x}_2)^T(F(\mathbf{x}_1) - F(\mathbf{x}_2)) < 0.$$

The proof is skipped due to space limitation.

Lemma 3. *Suppose C2 and C3 hold. Then $x_r^* > 0$ at equilibrium for all $r \in R$.*

Proof. We first show that there exists a $\overline{p} > 0$ such that $p_l^* \leq \overline{p}$ for $l \in L$. For any route $r \in s$, let

$$f_r(x_r) := \sup\{\phi_r(\mathbf{x}_s)\} \quad \text{s.t. } x_{-r} \in \mathbb{R}_+^{|s|-1}.$$

Note that

$$\frac{\partial \phi_r(\mathbf{x}_s)}{\partial x_r} = \mathbf{e}_r^T \frac{\partial \Phi_s}{\partial \mathbf{x}_s}\mathbf{e}_r < 0,$$

by C2, where $\mathbf{e}_r = (\ldots, 0, 1, 0, \ldots)$ and 1 is in the r_{th} entry. Thus $\phi_r(\hat{x}_r, x_{-r}) > \phi_r(\tilde{x}_r, x_{-r})$ if $\hat{x}_r < \tilde{x}_r$ for any x_{-r}, which means $f_r(\hat{x}_r) > f_r(\tilde{x}_r)$ if $\hat{x}_r < \tilde{x}_r$. In addition $f_r(x_r) < \infty$ if $x_r > 0$ by C3. Suppose there exists a $l_1 \in L$ with bandwidth c_{l_1} such that at equilibrium $p_{l_1}^* > \overline{p}$ for any $\overline{p} > 0$. Let $N_{l_1} := \sum_{r\in R} H_{l_1 r}$ and $\overline{p} = 2\max_{r:l\in r}\{f_r(\frac{c_{l_1}}{2N_{l_1}})\}$. Since $f_r(x_r)$ is a decreasing function of x_r and

$$f_r(x_r^*) \geq \phi_r(\mathbf{x}_s^*) = \frac{q_r^*}{2} > \frac{p_{l_1}}{2} \geq \frac{\overline{p}}{2} \geq f_r(\frac{c_{l_1}}{2N_{l_1}}),$$

we can obtain $x_r^* \leq \frac{c_{l_1}}{2N_{l_1}}$ for all $l_1 \in r$. However,

$$y_{l_1}^* = \sum_{r\in R} H_{lr}x_r^* \leq \frac{c_{l_1}}{2} < c_{l_1},$$

which means $p_{l_1}^* = 0$ by Eqn. (7). It contradicts $p_{l_1}^* > \overline{p}$. Thus there exists a $\overline{p}$ such that $p_l^* \leq \overline{p}$ for all $l \in L$, which means $q_r^* = \sum_{l\in r} p_l^*$ is also finite. Then if $x_r^* = 0$, we can obtain $\phi_r(0, x_{-r}^*) = \infty > q_r^*/2$, which contradicts C3. Thus, we have $x_r^* > 0$ for $r \in R$. □

Suppose there exist two equilibria $(\mathbf{x}_1, \mathbf{p}_1)$ and $(\mathbf{x}_2, \mathbf{p}_2)$ to Eqn. (3-4). Let $\delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ and $\delta\mathbf{p} = \mathbf{p}_1 - \mathbf{p}_2$, then for

$s \in S$,

$$\begin{aligned}
&\sum_{r\in s}(x_{r1}-x_{r2})\left(\phi_r(\mathbf{x}_{s1})-\frac{1}{2}q_{r1}\right)\\
=&\sum_{r\in s}\left(x_{r1}\left(\phi_r(\mathbf{x}_{s1})-\frac{1}{2}q_{r1}\right)-x_{r2}\left(\phi_r(\mathbf{x}_{s1})-\frac{1}{2}q_{r1}\right)\right)\\
=&-\sum_{r\in s}\left(x_{r2}\left(\phi_r(\mathbf{x}_{s1})-\frac{1}{2}q_{r1}\right)\right)\\
\geq&\ 0 \qquad (19)
\end{aligned}$$

where the second equality holds since $x_{r1}(\phi_r(\mathbf{x}_{s1})-q_{r1}/2)=0$ and the last inequality holds since $\phi_r(\mathbf{x}_{s1})-q_{r1}/2\leq 0$ by Eqn. (6). The following inequality can be obtained by similar manner.

$$\sum_{r\in s}(x_{r1}-x_{r2})\left(\phi_r(\mathbf{x}_{s2})-\frac{1}{2}q_{r2}\right)\leq 0 \qquad (20)$$

Subtract Eqn. (20) from Eqn. (19), we can obtain

$$\begin{aligned}
0\ \leq&\ \sum_{r\in s}\delta x_r\left(\phi_r(\mathbf{x}_{s1})-\phi_r(\mathbf{x}_{s2})-\frac{1}{2}\delta q_r\right)\\
=&\ \sum_{r\in s}\delta x_r\left(\phi_r(\mathbf{x}_{s1})-\phi_r(\mathbf{x}_{s2})\right)-\frac{1}{2}\delta\mathbf{x}_s^T\mathbf{q}_s\\
\leq&\ -\frac{1}{2}\delta\mathbf{x}_s^T\delta\mathbf{q}_s,
\end{aligned}$$

where the last inequality follows from Lemma 2 and C2 and equality can be obtained if and only if $\delta\mathbf{x}_s=0$. Sum the above inequality over $s\in S$,

$$0<-\sum_{s\in S}\delta\mathbf{x}_s^T\delta\mathbf{q}_s=-\delta\mathbf{x}^TH^T\delta\mathbf{p}. \qquad (21)$$

On the other hand, for $l\in L$,

$$\delta p_l(y_{l1}-c_l)\geq 0 \text{ and } \delta p_l(y_{l2}-c_l)\leq 0$$

which is obtained using similar manner as Eqn. (19). Then we can obtain

$$\delta p_l\delta y_l\geq 0$$

Sum over the above inequality over $l\in L$, we obtain

$$\delta\mathbf{p}^TH\delta\mathbf{x}\geq 0. \qquad (22)$$

which contradicts Eqn. (21) if $\mathbf{x}_1\neq\mathbf{x}_2$. Thus, $\mathbf{x}_1=\mathbf{x}_2$. Next we will show $\mathbf{p}_1=\mathbf{p}_2$. Since $x_r^*>0$ by Lemma 3, $\phi_r(\mathbf{x}_s)=q_r/2$ according to Eqn. (6). Let $\Phi(\mathbf{x}):=(\Phi_s(\mathbf{x}_s),s\in S)$, and we can get $H^T\mathbf{p}=\mathbf{q}=2\Phi(\mathbf{x})$. Then

$$H^T\mathbf{p}_1=\Phi(\mathbf{x}_1)=\Phi(\mathbf{x}_2)=H^T\mathbf{p}_2,$$

which means $\mathbf{p}_1=\mathbf{p}_2$ due to H has full row rank. Therefore there exists at most one equilibrium to Eqn. (3-4) under *C2-C3*.

A.3 Proof of Theorem 3

We borrow idea from contraction mapping analysis [9]. We will show that both $\dot{x}_r(t)$ and $\dot{p}_l(t)$ will be arbitrary small when t approach infinity and construct Lyapunov function based on that. Define $\delta\mathbf{x}:=\mathbf{x}-\mathbf{x}^\star$, $\delta\mathbf{p}:=\mathbf{p}-\mathbf{p}^\star$ as a perturbation of $\mathbf{x}$ and $\mathbf{p}$ around the equilibria. Let $\Lambda_k=diag\{k_r,r\in R\}$, $\Lambda_\gamma=diag\{\gamma_l,l\in L\}$. Then the Lyapunov function we develop is as follows:

$$V(\mathbf{x},\mathbf{p})=\delta\mathbf{x}^T\Lambda_k^{-1}\delta\mathbf{x}+\frac{1}{2}\delta\mathbf{p}^T\Lambda_\gamma^{-1}\delta\mathbf{p}. \qquad (23)$$

To show $\dot{V}(\mathbf{x},\mathbf{p})<0$, we need to show both $\delta\mathbf{x}^T\Lambda_k^{-1}(\mathbf{x})\delta\dot{\mathbf{x}}$ and $\delta\mathbf{p}^T\Lambda_\gamma^{-1}\delta\dot{\mathbf{p}}$ is smaller than 0 when $(\mathbf{x},\mathbf{p})\neq(\mathbf{x}^*,\mathbf{p}^*)$.

If $\delta\mathbf{x}\neq 0$, then

$$\begin{aligned}
&\delta\mathbf{x}^T\Lambda_k^{-1}\delta\dot{\mathbf{x}}\\
=&\ \sum_{r\in R}\delta x_r\left(\phi_r(\mathbf{x}_s)-\frac{q_r}{2}\right)_{x_r}^+\\
\leq&\ \sum_{r\in R}\delta x_r\left(\phi_r(\mathbf{x}_s)-\frac{q_r}{2}\right)\\
=&\ \sum_{r\in R}\delta x_r\left(\phi_r(\mathbf{x}_s)-\phi_r(\mathbf{x}_s^*)-\frac{\delta q_r}{2}\right)+\sum_{r\in R}\delta x_r\left(\phi_r(\mathbf{x}_s^*)-\frac{q_r^*}{2}\right)\\
\leq&\ \sum_{r\in R}\delta x_r\left(\phi_r(\mathbf{x}_s)-\phi_r(\mathbf{x}_s^*)\right)+\frac{1}{2}\sum_{r\in R}\delta x_r\delta q_r<-\frac{1}{2}\delta\mathbf{x}^TH^T\delta\mathbf{p}
\end{aligned}$$

The first inequality holds since $(\phi_r(\mathbf{x}_s)-\frac{q_r}{2})_{x_r}^+=\phi_r(\mathbf{x}_s)-\frac{q_r}{2}$ if $x_r>0$ and $\phi_r(\mathbf{x}_s)-\frac{q_r}{2}\leq 0$, $\delta x_r=-x_r^*$ if $x_r=0$. In the second inequality, we need

$$(x_r-x_r^*)\left(\phi_r(\mathbf{x}_s^*)-\frac{q_r^*}{2}\right)=x_r\left(\phi_r(\mathbf{x}_s^*)-\frac{q_r^*}{2}\right)\leq 0,$$

which holds since $x_r^*(\phi_r(\mathbf{x}_s^*)-\frac{q_r^*}{2})=0$ and $\phi_r(\mathbf{x}_s^*)-\frac{q_r^*}{2}\leq 0$ by the property of equilibrium to (3-4). The last inequality holds by Lemma 2 and C2.

On the other hand, using similar manner, we can obtain

$$\begin{aligned}
\delta\mathbf{p}^T\Lambda_\gamma^{-1}\delta\dot{\mathbf{p}}&=\sum_{l\in L}\delta p_l(y_l-c_l)_{p_l}^+\\
&\leq\sum_{l\in L}\delta p_l(y_l-c_l)\\
&\leq\sum_{l\in L}\delta p_l\delta y_l\\
&=\delta\mathbf{p}^TH\delta\mathbf{x}
\end{aligned}$$

Therefore if $\delta\mathbf{x}\neq 0$

$$\dot{V}(\mathbf{x},\mathbf{p})\ =2\delta\mathbf{x}^T\Lambda_k^{-1}\delta\dot{\mathbf{x}}+\delta\mathbf{p}^T\Lambda_\gamma^{-1}\delta\dot{\mathbf{p}}<0,$$

which means $\mathbf{x}(t)\to\mathbf{x}^*$. Recall that $\dot{p}_l=\gamma_l(y_l-c_l)_{p_l}^+$, it means $\dot{p}_l\to\gamma_l(y_l^*-c_l)_{p_l}^+=0$. By Theorem 2, there is a unique equilibrium to $(\dot{\mathbf{x}},\dot{\mathbf{p}})=0$ and it means $\mathbf{p}(t)\to\mathbf{p}^*$ or $\dot{\mathbf{p}}\not\to 0$. As a result, $V(\mathbf{x},\mathbf{p})$ is a Lyapunov function for the dynamical system (3-4).

A.4 Proof of Theorem 4

Assume the MP-TCP source runs algorithm

$$\phi_r(\mathbf{x}_s;\mu)=\mu\tilde{\phi}_r(\mathbf{x}_s)+(1-\mu)\hat{\phi}_r(\mathbf{x}_s)\quad \mu\in[0,1]$$

where algorithm M1 and M2 corresponds to $\mu=0$ and 1 respectively. Let x_g and τ_g be the throughput and RTT of the TCP NewReno source in the network of Fig. 1. Denote the inverse of $\frac{\partial\Phi_s(\mathbf{x}_s,\mu)}{\partial\mathbf{x}_s}$, whose existence is ensured by C2, by $D(\mu)$ and we have $\sum_{i\in s}D_{ij}(\mu)\leq 0$ by C4. The equilibrium is defined by $F(\mathbf{x},\mu)=0$ where $\mathbf{x}:=(\mathbf{x}_s,x_g)$ and F is given by:

$$\begin{aligned}
\Phi_s(\mathbf{x}_s,\mu)-\frac{1}{\tau_g^2x_g^2}\mathbf{1}&=0\\
\mathbf{1}^T\mathbf{x}_s+x_g&=c,
\end{aligned}$$

where the first equation follows from

$$\frac{p^*}{2}=\frac{1}{\tau_g^2x_g^2}=\phi_r(\mathbf{x}_s)\quad r\in s$$

and p^* is the congestion price at the bottleneck link. Applying implicit function theorem, we get

$$\begin{aligned}\frac{d\mathbf{x}}{d\mu} &= -\left(\frac{\partial F}{\partial \mathbf{x}}\right)^{-1}\frac{\partial F}{\partial \mu}\\ &= -\begin{bmatrix}\frac{\partial \Phi_s}{\partial \mathbf{x}_s} & \frac{2}{x_g^3}\mathbf{1}\\ \mathbf{1}^T & 1\end{bmatrix}^{-1}\begin{bmatrix}\tilde{\Phi}_s(\mathbf{x}_s)-\hat{\Phi}_s(\mathbf{x}_s)\\ 0\end{bmatrix},\end{aligned}$$

where the inverse exists by condition C2. Let

$$A := \frac{\partial \Phi_s}{\partial \mathbf{x}_s} - \frac{2}{x_g^3}\mathbf{1}\mathbf{1}^T \quad \text{and} \quad d := 1-\frac{2}{x_g^3}\sum_{i,j}D_{ij}(\mu)$$

Then

$$\begin{bmatrix}\frac{\partial \Phi_s}{\partial \mathbf{x}_s} & \frac{2}{x_p^3}\mathbf{1}\\ \mathbf{1}^T & 1\end{bmatrix}^{-1} = \begin{bmatrix}A^{-1} & -D\mathbf{1}d\\ -d\mathbf{1}^TA^{-1} & d^{-1}\end{bmatrix}.$$

Thus

$$\begin{aligned}\mathbf{1}^T\frac{\partial \mathbf{x}_s}{\partial \mu} &= -[\mathbf{1}^T 0]\left(\frac{\partial F}{\partial \mathbf{x}}\right)^{-1}\frac{\partial F}{\partial \mu}\\ &= -\mathbf{1}^TA^{-1}(\tilde{\Phi}_s(\mathbf{x}_s)-\hat{\Phi}_s(\mathbf{x}_s)). \qquad (24)\end{aligned}$$

By matrix inverse formula,

$$\begin{aligned}A^{-1} &= \left(\frac{\partial \Phi_s}{\partial \mathbf{x}_s}-\frac{2}{x_g^3}\mathbf{1}\mathbf{1}^T\right)^{-1}\\ &= D(\mu)+\frac{1}{\frac{x_g^3}{2}-\mathbf{1}^TD(\mu)\mathbf{1}}D(\mu)\mathbf{1}\mathbf{1}^TD(\mu),\end{aligned}$$

Substitute it into Eqn. (24), we have

$$\begin{aligned}&\mathbf{1}^TA^{-1}(\hat{\Phi}_s(\mathbf{x}_s)-\tilde{\Phi}_s(\mathbf{x}_s))\\ =&\left(1+\frac{\mathbf{1}^TD(\mu)\mathbf{1}}{\frac{x_g^3}{2}-\mathbf{1}^TD(\mu)\mathbf{1}}\right)\mathbf{1}^TD(\mu)(\tilde{\Phi}_s(\mathbf{x}_s)-\hat{\Phi}_s(\mathbf{x}_s))\\ =&\frac{x_g^3}{x_g^3-2\mathbf{1}^TD(\mu)\mathbf{1}}\sum_{r\in s}\left(\sum_{i\in s}D_{ir}(\mu)\right)(\tilde{\phi}_r(\mathbf{x}_s)-\hat{\phi}_r(\mathbf{x}_s))\\ \le&0.\end{aligned}$$

where the inequality follows from $D(\mu)$ is negative definite, $\sum_{i\in s}D_{ir}(\mu)<0$ and $\tilde{\phi}_r(\mathbf{x}_s)-\hat{\phi}_r(\mathbf{x}_s)\ge 0$. Thus, we have $\mathbf{1}^T\frac{\partial \mathbf{x}_s}{\partial \mu}\ge 0$ for $\mu\in[0,1]$, which means M2 will gain more throughput than M1.

A.5 Proof of Theorem 5

Let $h_r(\mathbf{x}_s):=\hat{\phi}_r(\mathbf{x}_s)-\tilde{\phi}_r(\mathbf{x}_s)$ and $H(\mathbf{x}_s):=(h_r(\mathbf{x}_s), r\in s)$, whose Jacobian, $\partial H/\partial \mathbf{x}_s=\partial\hat{\Phi}_s/\partial\mathbf{x}_s-\partial\tilde{\Phi}_s/\partial\mathbf{x}_s$, is negative definite by assumption. Now we only need to show $h_r(\mathbf{x}_s)\ge 0$, namely $\hat{\phi}_r(\mathbf{x}_s)-\tilde{\phi}_r(\mathbf{x}_s)\ge 0$. By condition C5, for any x_{-r}, we have

$$\begin{aligned}\lim_{x_r\to\infty}\hat{\phi}_r(\mathbf{x}_s)&\le\lim_{x_r\to\infty}\frac{1}{x_r^2\tau_r^2}=0\\ \lim_{x_r\to\infty}\hat{\phi}_k(\mathbf{x}_s)&\le\lim_{x_r\to\infty}\frac{1}{x_k^2\tau_k^2}<\infty \quad k\in s\setminus\{r\},\end{aligned}$$

which also holds for $\tilde{\Phi}_s(\mathbf{x}_s)$. Thus,

$$\begin{aligned}\lim_{x_r\to\infty}h_r(\mathbf{x}_s)&=\lim_{x_r\to\infty}(\hat{\phi}_r(\mathbf{x}_s)-\tilde{\phi}_r(\mathbf{x}_s))=0\\ \lim_{x_r\to\infty}h_k(\mathbf{x}_s)&<\infty \quad k\in s\setminus\{r\},\end{aligned}$$

By Lemma 2, for any $\mathbf{x}_s,\mathbf{z}_s\in\mathbb{R}_+^{|s|}$,

$$(\mathbf{z}_s-\mathbf{x}_s)^T(H(\mathbf{z}_s)-H(\mathbf{x}_s))<0,$$

which means

$$\begin{aligned}&(z_r-x_r)(h_r(\mathbf{z}_s)-h_r(\mathbf{x}_s)) \qquad (25)\\ <&-\sum_{k\in s,k\ne r}(z_k-x_k)(h_k(\mathbf{z}_s)-h_k(\mathbf{x}_s))\end{aligned}$$

Recall that $\lim_{x_r\to\infty}h_k(\mathbf{x}_s)<\infty$ for $k\ne r$, we have

$$A_r:=\limsup_{z_r\to\infty}|\sum_{k\in s,k\ne r}(z_k-x_k)(h_k(\mathbf{z}_s)-h_k(\mathbf{x}_s))|\le\infty.$$

Substitute it in Eqn. (25), we can obtain

$$\begin{aligned}&\limsup_{z_r\to\infty}(h_r(\mathbf{z}_s)-h_r(\mathbf{x}_s))\\ \le&\limsup_{z_r\to\infty}\frac{\sum_{k\in s,k\ne r}(z_k-x_k)(h_k(\mathbf{z}_s)-h_k(\mathbf{x}_s))}{z_r-x_r}\\ \le&\frac{A_r}{\lim_{z_r\to\infty}(z_r-x_r)}=0\end{aligned}$$

Thus for any $\epsilon>0$, we have

$$(h_r(\mathbf{z}_s)-h_r(\mathbf{x}_s))<\epsilon \text{ and } |h_r(\mathbf{z}_s)|\le\epsilon$$

provided z_r is large enough by condition C5. Then

$$h_r(\mathbf{x}_s)>h_r(\mathbf{z}_s)-\epsilon>-2\epsilon \quad \forall\epsilon>0,$$

which means $h_r(\mathbf{x}_s)\ge 0$ and verifies our claim that $\hat{\phi}_r(\mathbf{x}_s)\ge\tilde{\phi}_r(\mathbf{x}_s)$.

A.6 Proof of Theorem 6

We will show the results hold for any $n\in\mathbb{N}_+$. Since $\lim_{n\to\infty}\|x\|_n=\|x\|_\infty$, the results also holds for $n=\infty$.

Proof of $\phi_r(\mathbf{x}_s)$ satisfies C1 for $\beta\ge 0$

We will first show there exists a nonnegative solution $\mathbf{x}_s(\mathbf{q}_s)$ to Eqn. (6) for any $\mathbf{q}_s\ge 0$. Since

$$\{\mathbf{q}_s\mid q_r=\sum_{l\in L}H_{lr}p_l, r\in s,\mathbf{p}\ge 0\}\subseteq\mathbb{R}_+^{|s|},$$

it implies there exists a nonnegative solution to Eqn. (6) for any $\mathbf{p}\ge 0$. For simplicity, we relabel the subpaths of source s by $s=\{1,\dots,|s|\}$ such that $q_i\tau_i^2\le q_j\tau_j^2$ for $i\le j$. Let $\mathbf{z}_s:=\mathbf{x}_s/x_1$ and $C:=\|\mathbf{z}_s\|_n$. Then $\phi_r(\mathbf{x}_s)=\frac{q_r}{2}$ can be expressed as

$$1-\beta+\beta\frac{C}{z_r}=\frac{q_r\tau_r^2}{2}\|\mathbf{x}_s\|_1^2 \quad r\in s \qquad (26)$$

Divide the r_{th} equation to the 1_{st} equation and note that $z_1=1$, we obtain

$$z_r=\frac{q_1\tau_1^2\beta C}{(q_r\tau_r^2-q_1\tau_1^2)(1-\beta)+q_r\tau_r^2\beta C} \quad r\in s \qquad (27)$$

Note that $C=\|\mathbf{z}_s\|_n$ and substitute Eqn. (27) in it, we can obtain

$$\left(\sum_{r\in s}\left(\frac{q_1\tau_1^2\beta C}{(q_r\tau_r^2-q_1\tau_1^2)(1-\beta)+q_r\tau_r^2\beta C}\right)^n\right)^{\frac{1}{n}}=C$$

Let

$$\psi(C):=\left(\sum_{r\in s}\left(\frac{q_1\tau_1^2\beta C}{(q_r\tau_r^2-q_1\tau_1^2)(1-\beta)+q_r\tau_r^2\beta C}\right)^n\right)^{\frac{1}{n}}-C.$$

For any $C \geq 0$, we have

$$1 \leq \left(\sum_{r \in s} \left(\frac{q_1 \tau_1^2 \beta C}{(q_r \tau_r^2 - q_1 \tau_1^2)(1-\beta) + q_r \tau_r^2 \beta C} \right)^n \right)^{\frac{1}{n}} \leq |s|^{\frac{1}{n}}$$

Thus $\psi(1) > 0$ and $\psi(\infty) < 0$. Then there exists a $\tilde{C} \geq 1$ such that $\psi(\tilde{C}) = 0$. Plug $\tilde{C}$ back in Eqn. (27), we can obtain $\mathbf{z}_s$. Note that $\mathbf{z}_s \geq 0$ because $\tilde{C} \geq 1$. Plug $\mathbf{z}_s$ into Eqn. (26), we can solve $\|\mathbf{x}_s\|_1$ and $x_1 = \|\mathbf{x}_s\|_1 / \|\mathbf{z}_s\|_1$. Finally, we can calculate $\mathbf{x}_s$ by $\mathbf{x}_s = \mathbf{z}_s x_1$.

We skip the proof of $y_l^s(\mathbf{p})$ is a nonincreasing function of p_l since it is implied by that $\partial \Phi_s(\mathbf{x}_s)/\partial \mathbf{x}_s$ is negative semidefinite, which will be proved below. And it is easy to see $y_l^s(\mathbf{p}) \to 0$ as $p_l \to \infty$ so we skip it.

Proof of $\phi_r(\mathbf{x}_s)$ satisfies C2 and C3 for $\beta > 0$ and $|s| \leq 8$

Lemma 4. *Let $\mathbf{a} \in \mathbb{R}^n$ that satisfies $\sum_{i=1}^n a_i = 1$ and $\sum_{i=1}^n a_i^2 \leq 1$. Then for any n dimensional $\mathbf{z} \neq \mathbf{0}$,*

$$f(\mathbf{z}) := \sum_{i=1}^n z_i^2 + (\sum_{i=1}^n z_i)(\sum_{i=1}^n a_i z_i) > 0$$

provided $n \leq 8$.

Proof. Let $S_M := \{\mathbf{z} \mid \sum_{i=1}^n z_i = M\}$ and given any M, if we can show $f(\mathbf{z}) > 0$ for $\mathbf{z} \in S_M$, we can conclude that $f(\mathbf{z}) > 0$ since $\cup_{M \in \mathbb{R}} S_M = \mathbb{R}^n$. Given any M,

$$\min_{\mathbf{z} \in S_M} f(\mathbf{z}) = \min f(\mathbf{z}) \quad \text{s.t.} \ \sum_{i=1}^n z_i = M.$$

Its Lagrangian is given as

$$L(\mathbf{z}, \mu) = \sum_{i=1}^n z_i^2 + M(\sum_{i=1}^n a_i z_i) + \mu(\sum_{i=1}^n z_i - M),$$

where μ is the lagrangian multiplier. Let $\partial L / \partial z_i = 0$ for all $1 \leq i \leq n$ and substitute it back to $\sum_{i=1}^n z_i = M$, we can obtain $\mu = -3M/n$ and $z_i = \frac{M}{2}(\frac{3}{|s|} - a_i)$, which are the unique minimizer of $\min_{\mathbf{z} \in S_M} f(\mathbf{z})$. Then

$$\min_{\mathbf{z} \in S_M} f(\mathbf{z}) = \frac{M^2}{4} \left(\frac{9}{n} - \sum_{i=1}^n a_i^2 \right) \geq \frac{M^2}{4} \left(\frac{9}{n} - 1 \right)$$

When $M \neq 0$, $\min_{\mathbf{z} \in S_M} f(\mathbf{z}) > 0$ if $n < 9$. When $M = 0$, the minimizer is $\mathbf{z} = \mathbf{0}$ and $f(\mathbf{z}) > 0$ for $\mathbf{z} \in S_0 \setminus \{\mathbf{0}\}$. □

Lemma 5. *Given a $\mathbf{x} \in \mathbb{R}_+^n$, define a vector $\mathbf{a}$ in $\mathbb{R}^n$ as follows:*

$$a_i = \frac{2x_i}{\sum_{i=1}^n x_i} - \frac{x_i^p}{\sum_{i=1}^n x_i^p} \quad 1 \leq i \leq n$$

where $p \in \mathbb{N}_+$. Then $\sum_{i=1}^n a_i = 1$ and $\sum_{i=1}^n a_i^2 \leq 1$ for any integer $p \geq 1$.

Proof. It is straightforward to show $\sum_{i=1}^n a_i = 1$ and now we will show $\sum_{i=1}^n a_i^2 \leq 1$.

$$\begin{aligned}
\sum_{i=1}^n a_i^2 &= \frac{\sum_{i=1}^n x_i^{2p}}{\left(\sum_{i=1}^n x_i^p\right)^2} + \frac{4\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i\right)^2} - \frac{4\sum_{i=1}^n x_i^{p+1}}{\left(\sum_{i=1}^n x_i^p\right)\left(\sum_{i=1}^n x_i\right)} \\
&\leq 1 + \frac{4\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i\right)^2} - \frac{4\sum_{i=1}^n x_i^{p+1}}{\left(\sum_{i=1}^n x_i^p\right)\left(\sum_{i=1}^n x_i\right)} \\
&= 1 - 4\frac{\sum_{n \geq i > j \geq 1}(x_i - x_j)(x_i^{p-1} - x_j^{p-1})}{\left(\sum_{i=1}^n x_i\right)^2 \left(\sum_{i=1}^n x_i^p\right)} \\
&\leq 1
\end{aligned}$$

□

Now we begin to show the Jacobian $\partial \Phi_s(\mathbf{x}_s)/\partial \mathbf{x}_s$ is negative definite if $\beta > 0$ and negative semi-definite if $\beta = 0$. Let τ be the same round trip time for each route. Let $\Lambda_s = diag\{\mathbf{x}_s\}$ and

$$\mathbf{a}_s = \left(\frac{2x_r}{\|\mathbf{x}_s\|_1} - \frac{x_r^n}{\|\mathbf{x}_s\|_n^n}, r \in s \right).$$

Then the Jacobian of Φ_s at $\mathbf{x}_s$ can be written as

$$\frac{\partial \Phi_s}{\partial \mathbf{x}_s} = -\frac{2(1-\beta)}{\tau^2 \|\mathbf{x}_s\|_1^3} \mathbf{1}\mathbf{1}^T - \beta \frac{\|\mathbf{x}_s\|_n}{\tau^2 \|\mathbf{x}_s\|_1^2} \Lambda_s^{-1} \left(I_{|s|} + \mathbf{1}\mathbf{a}_s^T \right) \Lambda_s^{-1}$$

Next, we will show $I_{|s|} + \mathbf{1}\mathbf{a}_s^T$ is positive definite. For any $\mathbf{z}_s \in \mathbb{R}^{|s|}$,

$$\mathbf{z}_s^T (I_{|s|} - \mathbf{1}\mathbf{a}_s^T)\mathbf{z}_s = \|\mathbf{z}_s\|_2^2 - (\sum_{r \in s} z_r)(\sum_{r \in s} a_r z_r). \tag{28}$$

By Lemma 5, $\|\mathbf{a}_s\|^2 \leq 1$. Hence Eqn. (28) is always greater than 0 provided $|s| \leq 8$ by Lemma 4. Hence the Jacobian is negative definite if $|s| \leq 8$. For $\beta = 0$, the Jacobian degenerates to

$$\frac{\partial \Phi_s}{\partial \mathbf{x}_s} = -\frac{2}{\tau^2 \|\mathbf{x}_s\|_1^3} \mathbf{1}\mathbf{1}^T, \tag{29}$$

which is merely negative semi-definite.

The proof that $\phi_r(\mathbf{x}_s)$ satisfies C3 for $\beta > 0$ is straightforward so we skip it.

A.7 Proof of Lemma 1

For any $\lambda_1 \in \lambda(J)$, let $(\mathbf{z}_1, \mathbf{p}_1)$ be its corresponding eigenvector such that $\|\mathbf{z}_1\|_2 = \|\mathbf{p}_1\|_2 = 1$. Then we have

$$\lambda_1 \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{p}_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\Lambda_k & 0 \\ & \Lambda_\gamma \end{bmatrix} \begin{bmatrix} 2\frac{\partial \Phi}{\partial \mathbf{x}} & -H^T \\ H & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{p}_1 \end{bmatrix}$$

Premultiply $diag\{2\Lambda_k^{-1}, \Lambda_\gamma^{-1}\}$ at both sides, we get

$$\lambda_1 \begin{bmatrix} 2\Lambda_k^{-1} & 0 \\ & \Lambda_\gamma^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{p}_1 \end{bmatrix} = \begin{bmatrix} 2\frac{\partial \Phi}{\partial \mathbf{x}} & -H^T \\ H & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{p}_1 \end{bmatrix}$$

Now premultiply the conjugate of eigenvector $(\mathbf{z}_1, \mathbf{p}_1)$, we have

$$\begin{aligned}
\lambda_1 &= \frac{2\mathbf{z}_1^H \frac{\partial \Phi}{\partial \mathbf{x}} \mathbf{z}_1}{2\mathbf{z}_1^H \Lambda_k^{-1} \mathbf{z}_1 + \mathbf{p}_1^H \Lambda_\gamma^{-1} \mathbf{p}_1} \\
&= \frac{2}{2\mathbf{z}_1^H \Lambda_k^{-1} \mathbf{z}_1 + \mathbf{p}_1^H \Lambda_\gamma^{-1} \mathbf{p}_1} (\mathbf{Re}(\mathbf{z}_1^H \frac{\partial \Phi}{\partial \mathbf{x}} \mathbf{z}_1) + \mathbf{Im}(\mathbf{z}_1^H \frac{\partial \Phi}{\partial \mathbf{x}} \mathbf{z}_1)).
\end{aligned}$$

which means

$$\mathbf{Re}(\lambda_1) = \frac{2\mathbf{Re}(\mathbf{z}_1^H \frac{\partial \Phi}{\partial \mathbf{x}} \mathbf{z}_1)}{2\mathbf{z}_1^H \Lambda_k^{-1} \mathbf{z}_1 + \mathbf{p}_1^H \Lambda_\gamma^{-1} \mathbf{p}_1} = \frac{2\mathbf{z}_1^H (\frac{\partial \Phi}{\partial \mathbf{x}})^+ \mathbf{z}_1}{2\mathbf{z}_1^H \Lambda_k^{-1} \mathbf{z}_1 + \mathbf{p}_1^H \Lambda_\gamma^{-1} \mathbf{p}_1}$$

for any $\lambda_1 \in \lambda(J)$. Thus

$$\lambda_m(J) \leq \max_{(\mathbf{z},\mathbf{p}) \in S} \{ \frac{2\mathbf{z}^H (\frac{\partial \Phi}{\partial \mathbf{x}})^+ \mathbf{z}}{2\mathbf{z}^H \Lambda_k^{-1} \mathbf{z} + \mathbf{p}^H \Lambda_\gamma^{-1} \mathbf{p}} \}$$