In this paper, we examine the properties of game-theoretic objective function of individual sources that dictate their rate behavior. In our model, sources choose their own utility functions to maximize their throughput (utility proportional to throughput) while the network feedback (cost) function drives the system to a stable optimum point. The two facets of our motivation are to improve the steady-state packet loss characteristics and convergence of Kelly-type controllers. Earlier literatures studied packet loss penalty as a simply additive function; however in reality, the losses follow a product-form probability rule owing to the drops in the intermediate routers. We propose a novel *tangential controller* given a generic maximization function that proves to be a better rate adaptation algorithm for logarithmic utilities compared to the proportional controllers. We reconsider the additive packet loss penalty studied earlier and provide a penalty adjustment scaling factor. Our motivation here is to better understand the aggregate penalty loss characteristics across the path in the network. We establish the local stability criterion of our tangential controller through transfer function methods and give parametric bounds for increase/decrease gains. Finally, we implement a rate-based congestion control scheme named TTCP (tangential TCP) based on the tangential controller with a logarithmic cost function and develop an AQM-based explicit loss feedback scheme. We compare the relative performance of our controller with Kelly's proportional fairness scheme and observe that our controller has lower packet loss characteristics and faster convergence thus making it suitable for high bandwidth-delay product networks.

Index terms **– Mathematical programming/optimization, Control theory, Simulations, Utility optimization, Non-linear control**

I. INTRODUCTION

Providing performance incentives to end-to-end congestion control mechanisms can be one of the best ways to encourage deployment of behaving source agents in the Internet [1]. Recent studies have opened up several new avenues on optimizing flow control schemes using game-theoretic approaches [2], [3] and [4]. These have spurred a vast interest in deploying such end-to-end congestion schemes with varying user needs. There has been a significant shift in the paradigm among the researchers in analyzing congestion control techniques from traditional closed-loop flow control to game-theoretic optimization methods owing to its demonstration of success in the simulation and implementation. Many of these game-theoretic studies use analytical models to study the user rate behavior, given their utility and underlying network cost. One of the recent experiments based on this approach is Caltech's FAST TCP [12] that claims to provide throughput several times compared to TCP for high bandwidth-delay product networks. FAST TCP uses duality optimization theory to adjust sender's response based on both the queuing delay and packet loss as cost factors. The success of such strategic game theoretic-based congestion control schemes has been demonstrated through analytical study and simulations in several recent literatures as well [3], [5] and [7].

Utility-based techniques solve a problem of a particular nature such as improving the user response time (faster convergence) [2], [9], [13], [14], [16] and providing better fairness [4], [6], [15] to the users perceiving different utilities. Only a handful of literatures [2], [3] study the relative merits of use of one form of objective function and the rate adaptation against the other in terms of their properties such as stability and speed of convergence. Thus an analytical understanding based on optimization theories and basic calculus would be beneficial to enable the end applications to appropriately choose their utilities regardless of others in the network. From the network perspective, it is equally important to optimally make use of the link bandwidth given that users assign dissimilar utilities. In this paper, we analyze the form of objective functions and their characteristics against choosing the appropriate rate adaptation scheme. We propose a generic maximization function and establish the necessary criteria for convergence to efficiency and stability. For a given maximization function, we propose a novel *tangential controller* based on familiar tangent vector calculus that proves to be a better choice for rate adaptation scheme. We describe its origins and motivations and prove to have some interesting properties compared to the widely studied Kelly controller.

In order to study the characteristics of objective function, it is important to understand the form of network cost factor involved. Researchers have debated over the use of single to few bits ECN-style feedback against providing a fine-grain available bandwidth feedback as in XCP [11]. XCP develops a new congestion control scheme by introducing *precise congestion signal* to provide explicit feedback of the available bandwidth in the packet header. We are partly motivated by the design of XCP to allow explicit feedback on spare bandwidth available to flows to acquire fair share in quicker number of RTT steps than TCP. Our model assumes a continuous loss feedback (a function of network aggregate rate) of available bandwidth and all sources use the same network feedback to update their rates.

The rest of our paper is organized as follows: We present some of the related work in this area of research in the next section II. Section III describes our underlying model and problem motivation. Section IV introduces our novel *tangential controller* scheme motivated by the need to adjust loss penalty paid by earlier schemes. The tangential controller rate adaptation scheme uses a trajectory following technique to maximize a given objective function with the adjusted penalty. In section V, we establish the asymptotic stability of our controller using transfer function methods and give parametric bounds of increase and decrease constants. Finally, in section VI, we analyze and implement a rate-based *tangential TCP* (TTCP) scheme using NS-2 network simulator and show simulation results of the relative behavior with other forms of rate control schemes.

II. Related Work

Several forms of implicit network feedback are studied in the literature. Source agents infer the current network load through implicit means such as packet drop, queuing delay, asymptotic increase in queue sizes and variance of RTT, all of which are nonlinear in nature. As seen earlier, XCP [11] and ECN-based techniques develops explicit feedback mechanism. Kelly *et al*. [2] proposes a rate adjustment algorithm known as the dual form and uses the network loss rate feedback as pricing function. Traditional TCP and earlier utility-based studies [3], [7], [8], [9] assume loss feedback as a pricing function but recently La and Ananthram [16] and Alpcan and Basar [5] investigate the use of queuing delay as nonlinear pricing scheme and establishes the global stability of such controllers. An important challenge faced in nonlinear rate algorithm is to prove the existence of unique globally stable optimum point towards which all Liapunov trajectories converge. Asymptotic stability around equilibrium point is also desired for such systems.

Our model considers ECN-style explicit network feedback of cumulative packet losses across the link along the paths of the user. Cumulative packet losses are additive in nature along the path of the user as studied in current literatures [2] and [3]. Packet marking or dropping schemes however strictly follows a product form probability function $1-\prod_{l\in R}(1-p_l(\mathbf{x}))$, where $p_l(\mathbf{x})$ is the loss occurred at link l as a function of aggregate rate $\mathbf{x}$ through link l. In this paper, we study the additive penalty with a positive route-dependent price that the user pays and prove that this pricing scheme converges to a global optimum $\mathbf{r}^*$. Our tangential controller treats this additional penalty as an error-scaling factor.

Motivated by Mo and Warland [15], we study a generic objective function form for addressing the rationale behind fairness schemes, especially proportional and max-min fairness criteria. In [15], the authors show that if all users choose the same logarithmic homogeneous utility, the scheme converges to Kelly's proportional fairness. It is important to realize that the network allocates rates that turn out to be proportionally fair for homogeneous logarithmic utilities, whereas the user optimizes her/his throughput based on the price s/he pays through the cost function (utility minus the price for loss s/he pays). Thus, an open question remains whether there are any exceptions in the user rates for homogeneous logarithmic utility that does not result in proportional fairness. In later sections, we use constrained optimization theory to establish two fairness criteria that establishes a family of fairness schemes similar to proportional fairness.

Finally, the success of such strategic game theoretic-based congestion control schemes has been demonstrated through simulations in several recent literatures in [3], [5] and [7]. Kunniyur *et al*. [3] simulates agents with three utilities simultaneously sharing the links to prove the pronounced unfairness in their behavior when congestion occurs. Specifically, certain utilities with significant round-trip delays experience buffer starvation in FIFO queue with drop-tail mechanism, while certain utilities do not aggressively decrease window sizes when congestion is detected. Alpcan and Basar [5] develop a window-based TCP-friendly scheme with linear queuing delay as pricing scheme that proves the smoother convergence of flow rate as well as far less aggressive compared to TCP. Ganesh and Laevens [7] simulate rate adaptation scheme with a family of utility function (power law utility with different power of users' throughput) with heterogeneous price estimates and prove stability is not compromised.

III. Our Model

In this section, we present our basic analytical model for solving the optimization problem under link capacity constraints and provide its underlying motivation. Our underlying network model resembles that of Kelly *et al.* [2] but solves the inequality constraint problem using Kuhn-Tucker inequality conditions [10]. We start with a generic rate algorithm that quantifies the interesting properties of these controllers including stationarity, convergence and stability using simulations.

A. Our Model

Our analytical model considers an underlying network framework with a set of J resources or links $l \in J$ utilized by set of I users. Resources are links (or router queues) that have link capacities C_j and are capable of signaling end users by providing aggregate loss rate feedback across link l. Since we wish to explicitly evaluate the performance of our model with both end-to-end and AQM methods, we assume both forms of feedback control in our model and experiments.

User $i \in I$ in our model chooses utility function $U_i(x_i)$ which is strictly concave monotonically increasing function of rate x_i and is double differentiable over rate $x_k \geq 0$. For simplicity, we logically group the sources into sets such that the sources attach the same logical meaning to their utilities. Thus sources in the same set use homogeneous utilities. We show one such example in figure

Figure 1 below with sets S_1 and S_2. The model assumes a distributed approach in which bottleneck link feedback provides the aggregate feedback to sources in sets S_1 and S_2. The figure below shows our decentralized dual form optimization model with delayed aggregate price feedback to the users.

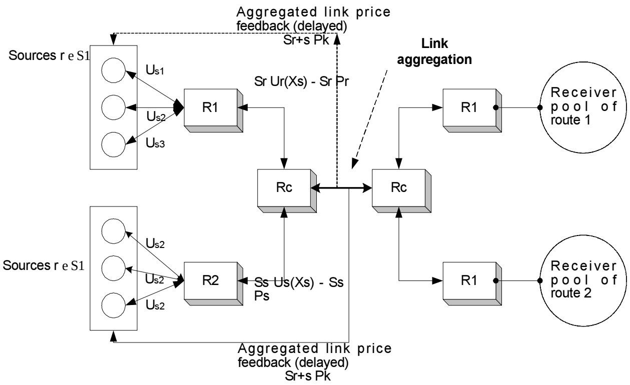

Figure 1 Closed-loop dual form feedback control system.

B. Background & Motivation

We are motivated by several recent literatures studying decentralized rate adaptation scheme in which the senders adjust the source rate at the same time maximizes the system objective function (utility minus cost). This led us to the following questions: 1) Can the rate adaptation algorithm be uniquely determined given the objective function to be maximized? That is, is there any systematic way for source k to adapt its rate x_k that "closely follows" the maximization function and that ultimately maximizes the system throughput of all users at a fair optimum rate x_k^* for all users k? 2) How does choosing of a rate control scheme affect its convergence to efficiency with minimum packet loss? 3) How can asymptotic and global stability be ensured of such a closed-loop feedback system under heterogeneous delayed conditions?

Literatures survey several forms of objective maximization functions but they do not necessary explain why the rate control behave the way it behaves. Specifically, they do not address the following problems:

¡

- What is the correlation between choosing a distributed rate adaptation scheme against maximizing a given objective?
- Are there other maximization functions that can converge to fair optimum rate and are Liapunov stable?
- For a given utility, what is the tradeoff between attaining the optimum throughput, speed of convergence to efficiency and compromise in stability with delayed feedback?
- Can packet loss scaling be made asymptotically sublinear with the increasing number of flows through a link?

¢

These questions serve as our primary motivation towards investigating controllers that may yield such desired properties.

C. Inequality Optimization Problem

Network transport agents are modeled as non-cooperating sender agents that maximize their own objective function. Objective functions are unbounded monotonically increasing utility minus a nonlinear penalty paid for the service. The domain of the objective functions or the feasible set of values is generally determined by equality or inequality constraints. Traditional techniques study the optimization problem with equality constraints with the primal and dual form algorithms [2]. Primal and dual optimization algorithms studied in [2], [9] treat the system maximization as equality constrained using Lagrangian formulation (looser sufficiency conditions). We believe that there are at least 2 problems in this approach. First, network problems always are known to operate under inequality conditions such as bottleneck bandwidth. Second, optimization theories suggest that inequality constraints with application of Kuhn-Tucker conditions [10] establish tighter bounds on shadow prices and hence we attempt to take this into account in our problem domain.

Our inequality problem formulation is as follows. Consider a constrained optimization problem for maximizing a given user objective function $U_k(x_k)$ that are strictly concave monotonically increasing function of throughput. User maximization function is constrained by inequality constraint for link $l, h_l(\mathbf{x}) \geq 0$, where $\mathbf{x}$ is the aggregate vector of all rates whose routes lie along link l. Assuming utilities are additive, the system-wide objective function $U(\mathbf{x})$ is a weighted sum of individual user utilities. We further assume that the feasible user allocation rates $x_k \geq 0$ are formed by the inequality constraint $h_l(\mathbf{x}) \geq 0$ is a closed and bounded set and hence a closed-ball D with optimum rate x_k^* as its radius. It is important to consider a closed-ball in order to establish a global optimum for our maximization function and to prevent Kuhn-Tucker conditions from failing at the global optimum (Kuhn-Tucker conditions are necessary but not sufficient for global maximum). Furthermore, it is important to realize that the closed-ball radius x_k^* is different for different flows until optimum fairness among all flows is established. However, there exists a global maximum of all flows for a given objective function: $x^* = \max\{x_k^*, \forall k\}$

Consider a network optimization problem that maximizes the system utility for all the users in the system. We assume our utilities are additive in nature constrained by the link capacity as shown below in equation.

$$\begin{cases} U(\mathbf{x}) = \max \sum_k U_k(x_k) \\ h(\mathbf{x}) = C - A^T\mathbf{x} \geq 0, x_k \geq 0, \forall k \end{cases} \tag{1}$$

The constraints in our optimization problem define the feasible set of rate allocation vectors that determine the optimum allocation rate $\mathbf{x}^*$ by solving the Lagrangian for the system of equations (1). Lagrangian for our system is defined by: $L(\mathbf{x}, \boldsymbol{\mu}) = U(\mathbf{x}) + \boldsymbol{\mu}^T h(\mathbf{x})$. Optimal solution $\mathbf{x}^*$ is determined by applying inequality theorem to the feasible set under the following conditions.

¡

- Condition 1: $\mu_l \geq 0, \ \ \mu_l h(x_k) = 0, \ \ h(x_k) \geq 0$, for all users k across link l.
- Condition 2: for all users k, we must have the following: ¢

$$\begin{cases} U_k{}'(x_k) + \mu_k \left.\dfrac{\partial h(\mathbf{x})}{\partial x_k}\right|_{x_k = x_k *} = 0 \\ x_k \geq 0 \end{cases} \tag{2}$$

An immediate observation from the above conditions is that the shadow price factor μ_l for link l is the same for all the users whose route passes through the link l. From condition 2 in (2) above, the first-derivative of the maximization function is given by equation (3) below.

$$U_k{}'(x_k*) = -\mu_l \frac{\partial}{\partial x_k}(C - \sum_m x_m) = \mu_l \tag{3}$$

We establish the following lemma 1 based on this.

£Users with homogeneous utilities $U_k(x_k)$ attain a fair share of the underlying link l with shadow price μ_l and the fair share for all users $\mathbf{x}^*$ is given by equation (4) below. ¦

$$\begin{cases} x_k^* & = & \dfrac{1}{U_k{}'(\mu_l)} \\ U_k{}'(x_k^*) & \neq & 0 \\ U_k{}'(\mu_l) & \neq & 0 \end{cases} \tag{4}$$

¨ We first note that condition 2 of equation (2) gives us the first-derivative of utility at optimum x_k^* as $U_k{}'(x_k^*) = \mu_l$. Assuming the shadow prices $\boldsymbol{\mu}$ are non-negative, we now apply the familiar inverse function theorem to obtain the first-derivative of the utility as a function of shadow price μ_l for link l at optimum x_k^*. Inverse function theorem is defined as follows. Given a continuous differential function $y = f(x)$ and a local optimum value $\mathbf{x}^*$, the inverse function $x = f^{-1}(y)$ exists near $\mathbf{x}^*$ if $f(\mathbf{x}^*) \neq 0$. In our case the function f happens to be $U'_k(x_k)$. Since the shadow price communicated to all users k (whose route lies along link l) is the same, their respective fairness share is the same for homogeneous utility functions. ■©

Since the shadow prices are chosen arbitrarily and since our first-derivative of objective function holds for all non-negative values of $\boldsymbol{\mu}$, we conclude that equation (4) defines the fairness among sources that have routes passing through link l. Indeed, in a single-link system, we find that this amounts to an exact fair share among all the users provided they use the same utility functions. It is this non-negativity condition on the shadow prices $\boldsymbol{\mu} \geq 0$ that enforces fairness among users and establishes tighter feasibility rate allocation sets.

We further extend the usefulness of the shadow prices and claim that it provides a family of fairness scheme, one of which is proportional fairness. The following lemma shows the validity of the claim.

£The optimization problem in equation (1) with additional fairness constraint $x_k^* - x_k \geq 0$ for all users k results in skewed fairness delay of $\frac{1}{x_i} - \frac{1}{x_j} = \mu_j - \mu_i$ for dissimilar users i and j sharing the same bottleneck link l. Proportional fairness is one such family of fairness scheme resulting from $\mu_i = \mu_j$.¦

¨ The Lagrangian of our new system includes the utility minus penalty constrained by additional fairness factor for each user k. This results in extra shadow prices λ_k for each user k. The Lagrangian of our new system may be written as follows.

$$L(\mathbf{x}, \boldsymbol{\mu}) = U(\mathbf{x}) - \sum_l p_l(\mathbf{x}) + \boldsymbol{\mu}^T h(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{x}^* - \mathbf{x}) \tag{5}$$

Applying condition 2 in equation (2) and simplifying the resulting expression yields skewed fairness delay factor for 2 users i and j that share the same underlying bottleneck link l resulting in equation (6) below.

$$\frac{1}{x_i} - \frac{1}{x_j} = \mu_j - \mu_i \tag{6}$$

■©

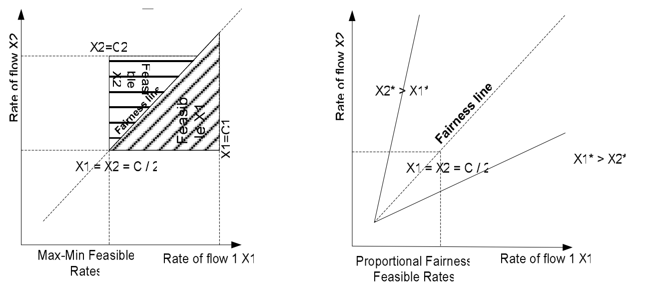

Figure 2 Feasible Rate Allocations for Max-Min and Proportional Fairness schemes for 2-flow system

We show the results of lemma 2 in figure

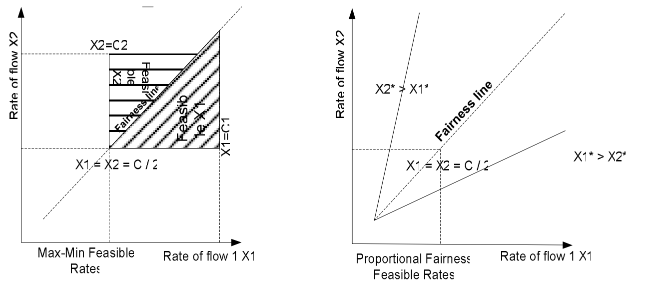

Figure 2 above that compares 2 feasible allocation rates sets for max-min and proportional fairness schemes for 2 flows. Max-min sets can take values in set $C_1/2 \le X_2 \le C_1$ and $C_2/2 \le X_2 \le C_2$, bounded by the right-triangles shown above. On the other hand, proportional fairness are defined by $\sum_{k \in R} \frac{x_k^* - x_k}{x_k} \le 0$ and hence can widely vary between the two lines shown $X_1^* > X_2^*$ and $X_1^* > X_2^*$. The result in lemma 2 shows precisely this.

IV. Tangential Controller

The inequality optimization problem described above naturally leads us to develop our novel *tangential controller*. In this section, we develop the necessary theory to demonstrate the design of our controller. First, we introduce a trajectory-following technique and prove it is indeed related to Kuhn-Tucker inequality optimization. Second, we show the need for additive (positive) packet loss penalty to scale the losses aggressively. Third, we utilize this (positive) packet loss scaling factor and theorize that our pricing scheme with additional loss scaling factor results in much smaller error in source rate evolution.

In modeling the controller, we attempt to address some of the motivations described in earlier section II for a decentralized rate adaptation scheme. Specifically, we investigate the correlation between a rate adaptation scheme and given objective maximization function. We design a novel trajectory-following technique that uniquely maximizes the objective function at the finite optimum rate. Using logarithmic utility, we contrast our scheme with Kelly-style proportional controller. We find that our scheme converges faster and much closer to the bottleneck bandwidth with up to 4 times lesser relative packet loss. The trajectory following hypothesis is proved using inequality optimization problem in which the user pays an additional (positive) packet loss penalty in addition to the current penalty paid, to converge to an optimum. The additional penalty is supported by our underlying theory that packet losses across all the links in the user's route are non-additive in nature and hence an appropriate error scaling factor is required. Indeed, cumulative packet drops across droptail-enabled routers (in a parking lot topology) results in a product-form probability given by $1 - \prod_{l \in R}(1 - p_l(x_l))$. Cumulative losses occur across all the links $l \in R$ along path R for the user.

While the product-form loss penalty function may be a suitable for adjusting our user rates, it is only be applicable only to specific topologies (such as parking lot) and it invariably assumes droptail queues across links in the user route. Routes with heterogeneous queuing along the paths may not necessary take the product-form penalty function. In contrast, we believe a general-form additive penalty function with suitable *loss scaling* factor can add to our knowledge of the form of packet loss penalty. Thus, we have the following motivations to study this form of additive penalty function with suitable scaling factor:

¡

- What positive error scaling factor is required and how should it be calculated to design a rate adaptation scheme that converges to optimum?
- How aggressive can the scaling penalty be while maintaining a stable rate control equation?
- Can the scaling factor be utilized to bring the system optimum rate to lesser than the bottleneck capacity?
- Does the user rate control equation tolerate negative packet loss penalty?

¢

We justify the addition of an error scaling factor i.e. points 1 and 2 in section B and C below. Our later section establishes point 3 using simulations to prove that the efficiency point of our controller is closer to bottleneck capacity. We leave point 4 for further study.

A. Trajectory-Following Algorithm

Our trajectory-following formulation is as follows. Consider a positive-definite source rate that evolves according to the first-order differential equation shown in equation (7) below.

$$\frac{d\mathbf{r}}{dt} = \frac{d\varphi(\mathbf{r}, \eta)}{dt} = f(\mathbf{r}) \tag{7}$$

The source rate vector $\mathbf{r}$ is adjusted according to trajectory tracking function $f(\mathbf{r})$ that is yet to be determined. Functions $\varphi(\mathbf{r},\eta)$ are said to be *flows* that are solutions of the rate differential equation such that every constant η yields an integral curve of equation(7). We claim that there exists at least one such integral curve of (7) that starts at a non-zero local minimum $\mathbf{r_0}$ (initial source rate) and uniquely maximizes to $\mathbf{r}^*$ along the given objective function curve $f(\mathbf{r})$.

£Suppose $G(\mathbf{r})$ is the objective maximization function. The tangent vector at any point $\mathbf{r}$ (gradient at $\mathbf{r}$) of the cost maximization function $G(\mathbf{r})$ at every step yields the closest possible trajectory towards unique maximum $\mathbf{r}^*$.¦

¨Consider any feasible allocation vector $\mathbf{r} \in D$, the closed ball of allocation rates. If our incremental rate change is still bounded by ball D, $\mathbf{r} + \Delta\mathbf{r} \in D$, then we can define our objective function around the neighborhood of $G(\mathbf{r})$ using Taylor series expansion given in (8) below.

$$
\begin{aligned}
G(\mathbf{r}+\Delta\mathbf{r}) &\geq G(\mathbf{r}), & \forall \mathbf{r} \\
G(\mathbf{r}) + \frac{\partial G(\mathbf{r})}{\partial \mathbf{r}}\Delta\mathbf{r} + R(\Delta\mathbf{r}) &\geq G(\mathbf{r}), & \forall \mathbf{r} \\
\Delta\mathbf{r}\left(\frac{\partial G(\mathbf{r})}{\partial \mathbf{r}} + \frac{R(\Delta\mathbf{r})}{\Delta\mathbf{r}}\right) &\geq 0, & \forall \mathbf{r}
\end{aligned} \tag{8}
$$

Noticing that $\Delta\mathbf{r}$ is positive and $\lim_{\Delta\mathbf{r}\to 0}\frac{\mathrm{R}(\Delta\mathbf{r})}{\Delta\mathbf{r}} = 0$, the gradient of the objective function is positive:

$$\frac{\partial G(\mathbf{r})}{\partial \mathbf{r}} \geq 0\,. \tag{9}$$

The objective function $G(\mathbf{r})$ grows monotonically increasing with time and because of this property, an integral curve solution for (7) is given by equation (10) below.

$$
\begin{aligned}
\frac{d\phi(\mathbf{r},\eta)}{dt} &= \left[\frac{\partial G(\mathbf{r})}{\partial \mathbf{r}}\right]^T \\
\frac{dG(\mathbf{r})}{dt} &= \frac{\partial G(\mathbf{r})}{\partial \mathbf{r}}\frac{d\phi(\mathbf{r},\eta)}{dt} \geq 0, \text{ positive definite}
\end{aligned} \tag{10}
$$

One such proof of the tangent vector and solution (10) is given in [17] in nonlinear constrained optimization. Since the source rate evolves "closely" according to the trajectory followed along the gradient of the cost optimization function $G(\mathbf{r})$, we term our controller a *tangential controller.* ■©

£The gradient of objective maximization function $G(\mathbf{r})$ is concave and evolves strictly along the *direction* of the gradient $\frac{\partial h(\mathbf{r})}{\partial \mathbf{r}}$ of the inequality constraint $h(\mathbf{r})$. ¦

¨ The theorem is proved using Farkas's lemma [17] whose geometric interpretation is that the steepest increase in the gradient $\frac{\partial G(\mathbf{r})}{\partial \mathbf{r}}$ must lie along the direction of the constraint gradient $\frac{\partial h(\mathbf{r})}{\partial \mathbf{r}}$, which is negative. Condition 2 in equation (2) of Kuhn-Tucker suggests the same that our constraint gradient is non-positive due to the positive shadow prices and hence the gradient of development indeed satisfies Farkas's lemma. Our immediate observation is that the tangent vector scheme is related to the Kuhn-Tucker inequality conditions. ■©

We illustrate the trajectory-tracking mechanism using the figure Figure 3 below. The simple linear constraint $h(\mathbf{r})$ is shown by thick linearly decreasing function $h(\mathbf{r}) = C - \mathbf{r}$, where C is the bottleneck capacity. As long as the constraint is fulfilled, the objective function monotonically increases along the curve $G(\mathbf{r})$. At the critical rate $\mathbf{r}^*$, the objective function settles at a constant gradient and hence the rate becomes steady state. The downward directional vector indicates the negative constraint gradient into taking effect at optimum $\mathbf{r}^*$.

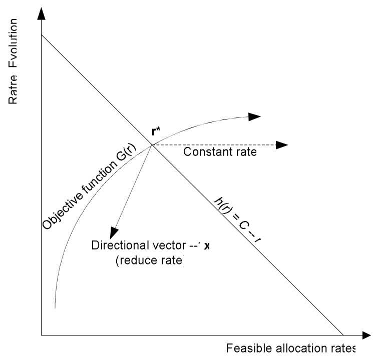

Figure 3 Trajectory-tracking technique and linear constraint

B. Packet Loss Penalty

Our motivation to reconsider packet loss penalty functions arises from two main sources. Existing literatures [2], [3] consider additive packet losses across the links along the path of the user. Ganesh *et al.* [7] consider an iso-elastic exponentially weighted moving average price estimator with the goal of keeping the link utilization close to the bottleneck capacity. A similar gradient-projection price estimator was developed by

Low and Lapsley [9] in which link prices are adjusted in the opposite direction to the gradient of the price at every step. Their controller adjusts the link prices according to the aggregate price across the bottleneck link. The price gradient is the gradient of the dual objective function (the Lagrangian itself).

Our penalty adjustment is similar to the controller developed by Low and Lapsley [9] but we develop our motivation from optimization theory. The objective of their price adjustment controller is to solve for the source rate as a function of optimum price. Our tangential controller algorithm solves the dual form distributed source rate introduced by Kelly *et al.* [2]. We use sequential nonlinear programming [17] technique that adds penalty functions to the maximization function such that successive iterations lead to convergence of the given sequence. The goal is to choose an appropriate penalty function that converges to an optimum. Using the tangent vector controller developed earlier, we show that our pricing scheme indeed converges to a unique optimum.

In our model, link l has packet loss $p_l(\mathbf{r})$ plus an additional route-dependent penalty scaling factor $Q_{R_k}(\mathbf{r})$. The scaling factor is path-dependent such that R_k represents the path for user k. Thus the net cost paid by the user k in our pricing scheme is given by equation (11) below.

$$W_k(r_k) = \alpha U_k(r_k) - \beta\left(\sum_{l\in R}\int_0^{\sum x_s} p_l(x)dx - \beta_c Q_{R_k}(\mathbf{r})\right)$$
$$Q_{R_k}(\mathbf{r}) = \max\left(\sum_{s\in l} r_s - C_l\right)^2, \forall l \in R_k \tag{11}$$

In contrast, we compare our equation (11) with the original pricing scheme in equation (12) as studied by Kelly *et al.* [2].

$$M_k(r_k) = U_k(r_k) - \beta\sum_{l\in R}\int_0^{\sum x_s} p_l(x)dx \tag{12}$$

£Path-dependent penalty scaling factor $Q_{R_k}(\mathbf{r})$ solves for the optimum rate $\mathbf{r}^*$ for the pricing scheme $\mathbf{W}(\mathbf{r})$. ¦

¨The above theorem requires that the scaling factor $Q_{R_k}(\mathbf{r})$ be bounded and leads to the convergence of the sequence. The scaling factor introduced is the square of the constraint $h_k(r_k)$ and this is borrowed from the sequential nonlinear programming method. The iterative method introduces large penalties initially and when the rate differs from the constraint widely, due to this scaling factor. Thus as the source rate stays closer to the bottleneck bandwidth, the penalty is driven to zero leading to the natural form (12). The penalty function is chosen to be the maximum of all penalties across the user's path. This takes into effect the addition of newer flows along the path while the bottleneck link itself changes.■ ©

£Our pricing scheme $\mathbf{W}(\mathbf{r})$ is bounded and results in a convergent sequence in successive iterations. ¦

¨We prove the convergence of this lemma using the tangent rate adaptation scheme in lemma 8 in section C below. ■©

In order to demonstrate that our penalty converges to a steady state, we perform experiments that compare our tangential controller with proportional fairness. We perform 3 experiments with the given 3-flow topology below and show its significance.

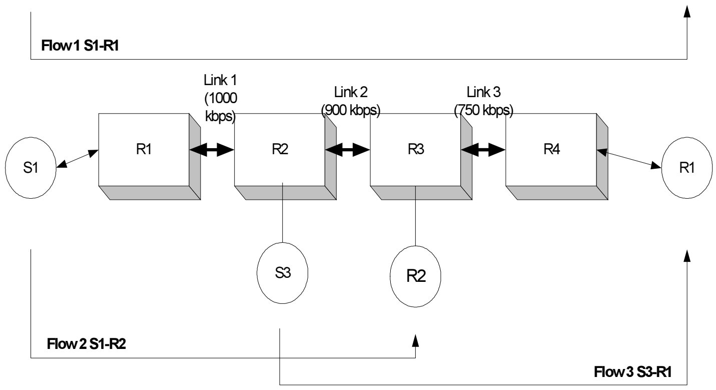

Figure 4 Parking lot topology with 3 flows

Figure Figure 4 above shows the parking-lot topology for our experiments. We consider 2 sources, 2 receivers and 3 bottleneck links with 3 flows S_1-R_1, S_1-R_2, and S_3-R_1. Flows either use proportional or tangential controllers for our experiments.

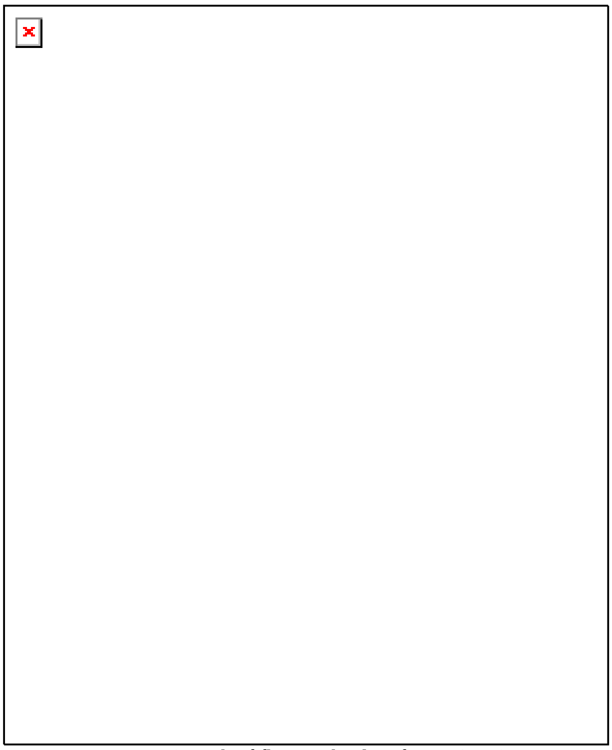

Figure 5 Growth of flow packet loss for PF / TTCP

In our experiments, we compare the cumulative flow losses of each of the 3 flows S_1-R_1, S_1-R_2, S_3-R_1 and across links 1, 2 and 3. Figures Figure 5a and figure Figure 5b show the cumulative additive and multiplicative losses for proportional controllers and figure Figure 5c shows cumulative losses for our tangential controller. All flows start at initial rate α=10kbps with β=0.5. We find that additive losses were consistently smaller compared to multiplicative for proportional controller. The cumulative losses of flows of tangential is roughly about twice that of additive proportional suggesting less aggressive utilization. On the other hand, the individual link losses are up to 4 times smaller than that of the proportional (not shown above).

C. Pricing-Scheme and Rate Adaptation

In this subsection, we demonstrate that our network pricing scheme establishes a unique equilibrium **r*** between the user paid price and the network allocated rates. Our previous subsection describes the necessary motivation for additional packet loss penalty. In this subsection, we use the additional penalty to design a pricing scheme. We find that all users sharing a common link *l* pay an additional penalty that leads to faster convergence to optimum point and reduced overall packet loss for all the flows through link *l*. Our aim now is to develop a network rate allocation scheme P with dependence on the following parameters shown in equation (13) below.

$$\begin{aligned} &\text{P: } NETWORK(\mathbf{U}(\mathbf{r}), h(\mathbf{r}), p_l(\mathbf{r}), \mathbf{Q}(\mathbf{r})) \\ &\mathbf{U}(\mathbf{r}) = \left[U_1(r_1)\ \ U_2(r_2) \ldots U_k(r_k)\right] \\ &\mathbf{Q}(\mathbf{r}) = \left[Q_{R_1}(\mathbf{r})\ \ Q_{R_2}(\mathbf{r}) \ldots Q_{R_k}(\mathbf{r})\right] \end{aligned} \tag{13}$$

Utilities are user-dependent strictly concave increasing functions of user's throughput r_k constrained by $h(\mathbf{r})$. We now claim that our network problem P in equation (13) is solved by our novel pricing scheme (11) with scaled penalty. We theorize that our pricing scheme indeed establishes a unique optimum **r*** only if the source rate adaptation adopts the tangent-vector algorithm. The gradient vector algorithm simply requires the rate differential equation for user *k* to vary according to the gradient of objective function with respect to the rate r_k. This is given by:

$$\frac{dr_k}{dt} = r_k \frac{\partial M_k(r_k)}{\partial r_k} - \beta\beta_c \frac{\partial Q_k(r_k)}{\partial r_k}. \tag{14}$$

Expanding the right side of (14), we get the following:

$$\frac{dr_k}{dt} = r_k \left(\alpha U'_k(r_k) - \beta \sum_{l\in R} p_l(x)\right) - \beta\beta_c \frac{\partial Q_{R_k}}{\partial r_k}. \tag{15}$$

£With a strictly concave increasing utility $\mathbf{U}(\mathbf{r})$, our objective maximization (11) forms the Liapunov for our source rate control equation (14). This solves our optimization problem P by maximizing at the unique optimum **r***.¦

¨For a strictly concave increasing utility function $\mathbf{U}(\mathbf{r})$, we observe the rate change of our Liapunov $W_k(r_k)$ as given in the equation (16) below.

$$\begin{aligned} \frac{d\mathbf{W}(\mathbf{r})}{dt} &= \sum_k \frac{\partial W_k(r_k)}{\partial r_k} \frac{dr_k}{dt} = \\ &= \sum_k \left(\frac{\partial M_k(r_k)}{\partial r_k} + \beta\beta_c \frac{\partial Q_k(r_k)}{\partial r_k} \right) \frac{dr_k}{dt} \end{aligned} \tag{16}$$

For the given rate adaptation algorithm given in (14), the maximization is positive definite as show below.

$$\begin{aligned} &\frac{d\mathbf{W}(\mathbf{r})}{dt} = \\ &= \sum_k \left(\begin{aligned} &r_k \left(\frac{\partial M_k(r_k)}{\partial r_k}\right)^2 - \beta\beta_c \left(\frac{\partial Q_k(r_k)}{\partial r_k}\right)^2 + \\ &+(r_k - 1)\frac{\partial M_k(r_k)}{\partial r_k}\frac{\partial Q_k(r_k)}{\partial r_k} \end{aligned} \right) \geq 0 \end{aligned} \tag{17}$$

The time derivative of cost optimization function is positive definite because the rate partial differential of functions $M_k(r_k)$ and $Q_k(r_k)$ are strictly positive as shown below.

$$\begin{aligned} &\frac{d\mathbf{W}(\mathbf{r})}{dt} \geq 0 \\ &\because \frac{\partial M_k(r_k)}{\partial r_k} \geq 0 \forall r_k \geq 0 \\ &\because \frac{\partial Q_k(r_k)}{\partial r_k} \geq 0 \forall r_k \geq 0 \end{aligned} \tag{18}$$

Thus, the monotonically increasing Liapunov gradient with respect to time uniquely maximizes the rate evolution to optimum **r***. As shown in Figure 3, the constraint $h(\mathbf{r})$ takes effect as the rate approaches the bottleneck bandwidth and it is at this point that the Liapunov maximizes. ■©

£Our scaled pricing scheme in (11) introduces lesser error factor along the rate trajectory compared to the original pricing scheme (12).¦

¨In order to prove the error factor, we calculate the cumulative area of the error curve along the trajectory starting at the initial rate $\mathbf{r}_0$ to optimum **r***.

$$\begin{aligned} E_{W_k}(r_k,\eta) &= \int_{r_{k0}}^{r_k^*} \left(W_k(r_k) - \eta r_{W_k}\right) dr_{W_k} \\ E_{M_k}(r_k,\eta) &= \int_{r_{k0}}^{r_k^*} \left(M_k(r_k) - \eta r_{M_k}\right) dr_{M_k} \end{aligned} \tag{19}$$

We claim that the error resulting by our pricing scheme (11) is lesser than the originally used one (12). That is, error condition $|E_{W_k}|<|E_{M_k}|$. Evaluating integral (19) along the curve and simplifying $|E_{W_k}|-|E_{M_k}|$ yields the following equation (20).

$$\begin{aligned} &E_{W_k} - E_{M_k} = \frac{\beta\beta_c}{3}\left(3r_k C - \left(r_k^*\right)^2 - 3C^2\right) \\ &E_{W_k} - E_{M_k} \leq 0 \text{ as } \lim r_k \to C \end{aligned} \tag{20}$$

This establishes our condition that for $|E_{W_k}|-|E_{M_k}|$, we must have $\beta_c > 0$. Notice that if $\beta_c=0$, our controller is equivalent to the well-studied proportional controller.■©

V. Stability Analysis

In this section, we establish the local stability of our controller with homogeneous delays using fluid-flow approximation and transfer function models. Using the Jacobian linearization around the equilibrium point, we study the tolerance to perturbation and prove that our controller indeed has a high phase margin. For a given finite time delay, we prove that the open loop transfer function of our plant controller does not encircle negative unity only if our decrease parameters are bounded, $\beta_c > 0$ and $0 < \beta(1+\beta_c)De^{\alpha T} < \frac{(2n+1)}{2}\pi,\ n \geq 0$. We thus establish that our controller is delay-tolerant as long as the delay is finite.

A. Conditions for Local Stability

Consider the generic model depicted in figure

Figure 1 where the sources in sets S_1 and S_2 have an associated route to its unique receiver in the receiver pool. We consider constant homogeneous round-trip total delay of *D* of all sources in this set. Thus we assume that queuing delays are negligible and we take only the propagation delay into account. We now study the general case state-space analysis of N-flows with homogeneous round-trip delays. To begin with, consider a simple model consisting of single bottleneck link with finite N but arbitrary number of flows across the link. Such a model can be easily extended to include more bottleneck links.

Consider the rate transfer vector **X** of source rates whose open-loop system matrix is defined as $\mathbf{G}(s)$. The open-loop vector is a function of state matrix **A**, delayed state-matrix $\mathbf{A_d}$, input **B** and input-output matrix **C**. The open-loop transfer function of our system with multiple-input and multiple-output is given by equation below.

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A} - \mathbf{A_d}\mathbf{D})^{-1}\mathbf{B}\,. \tag{21}$$

Linearizing our rate controller (15) around the optimum point and taking the Laplacian results in the following form [18].

$$\begin{aligned} s\mathbf{X}(s) &= \mathbf{AX} + \mathbf{A_d}\mathbf{DX} + \mathbf{BU} \\ \mathbf{Y}(s) &= \mathbf{CX} \end{aligned}\,. \tag{22}$$

The open-loop transfer function $\mathbf{G}(s)$ has the same elements across all rows and columns owing to nature of the symmetric circulant matrix $\mathbf{M}(s) = (s\mathbf{I} - \mathbf{A} - \mathbf{A_d}\mathbf{D})^{-1}$ and each element is given as below [18].

$$\begin{aligned} G_{ik}(s) = &\frac{s + \alpha' - \beta'' e^{-sD}}{(s + \alpha' + (N-3)\beta'' e^{-sD})} * \\ &* \frac{1}{(s + \alpha' + \beta'' e^{-sD})^{n-1}}, \forall i,k \end{aligned} \tag{23}$$

In this equation, we have the following definitions.

$$\begin{aligned} \alpha' &= \frac{n\alpha\beta(1+\beta_c)}{C\beta(1+\beta_c) + n\alpha} \\ \beta'' &= \frac{C\beta^2(1+\beta_c)^2}{n(C\beta(1+\beta_c) + n\alpha)} \end{aligned} \tag{24}$$

Equation (22) also defines an additional delay matrix $\mathbf{D}=diag\{e^{-sD}\}$ of equal delay *D* in the system.

£ Consider the closed-loop feedback system with loop transfer function (21). Consider a single bottleneck link *l* with *N*-flows and each flow having an optimum source rate r^* with a non-zero positive delay *D*. Then, the system is locally asymptotically stable if the following bound holds. ¦

$$0<\beta(1+\beta_c)De^{\alpha' T}<\frac{(2n+1)}{2}\pi,\ \ n\geq 0 \tag{25}$$

¨We notice that the characteristic polynomial of our open-loop system transfer function is the determinant of circulant matrix **M** given by the following characteristic polynomial for $N>1$ flows [18].

$$\begin{aligned}&(s+\alpha'+(N-3)\beta(1+\beta_c)e^{-sD})*\\&*(s+\alpha'+\beta(1+\beta_c)e^{-sD})^{n-1}=0\end{aligned} \tag{26}$$

Asymptotic Nyquist stability requires that the roots of our characteristic equation be lesser than unity. In our case, we argue that our polynomial $s+\alpha'+\beta(1+\beta_c)e^{-sD}$ has negative real roots resulting in the stability our controller (15). The roots of this polynomial are given by Lambert W function [19] and the only negative range of values for which our polynomial holds is given by equation (25) defined above. We plot the frequency response of this polynomial that makes us the open-loop transfer function in the following figure 6 Bode diagram for three different values of delay *D*. ■©

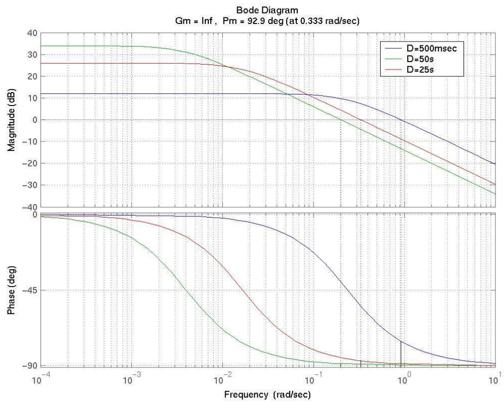

Figure 6 Bode plot of open-loop transfer function for various delays

VI. SIMULATIONS

In this section, we perform network simulator NS-2 based experiments to verify our theoretical results. We develop discretized delay-tolerant source and sink agents that emulate our rate control behavior using end-to-end and explicit AQM loss feedback models. We attempt to evaluate the stationarity and convergence properties of logarithmic proportional controller with additive penalty across the bottleneck links against our penalty scaling factor using tangential rate controller. All controllers adjust their rates based on explicit end-to-end or AQM loss feedback from the bottleneck links available in the packet headers. We perform experiments using with three explicit loss feedback AQM schemes including max-min, proportional fairness and tangential AQM-based loss adjustments and one form of sliding average packet end-to-end packet loss. Experimental results show that our end-to-end logarithmic proportional controller suffers from extrapolated estimation of rates of other flows and hence the bottleneck the large deviation in estimating the available bandwidth. A sliding-window average rate calculation or exponential weighted average requires estimating of sliding loss to emulate a smoother router-based feedback. Our observation is that extrapolating the aggregate rate or the averaged loss estimation leads to AIMD-type oscillations. Since one of our motivations is keep our steady-state oscillations closer to the bottleneck bandwidth, we investigate on developing AQM-based schemes.

We develop rate-based *tangential* source and sink agents that emulate our rate control behavior using delayed feedback model. Our stabilized source takes history-based rate adjustment owing to the roundtrip-delay. The granularity of aggregate link loss calculation in our AQM scheme is fine-tunable and the sources respond to the newer loss calculation only once during this AQM interval. These two factors stabilize the sources for an arbitrary number of flows and for chosen constants α and β. In our simulation experiments, we consider a standard parking-lot topology with 3 flows with 2 intermediate bottleneck link of our tangential AQM scheme with link capacity 500kbps. All sources start at the same interval and adjust with initial rate of 20kbps, and with constants α=20kbps, and β=0.5. We compare the link loss characteristics of max-min and proportional controller in this setup with their flow rate evolution. Max-min fairness results when our AQM scheme updates the packet header with the highest packet loss of the most congested link across the path from the source, whereas proportional penalty adds the link prices along all the queues in the path. In addition to the proportional loss, our proposed tangential source requires an AQM loss scaling factor inserted in the ACK packet header in the return path from the receiver.

A. Max-Min and Proportional AQM Feedback

We observe the relative performance of max-min and proportional controllers with sample 3 flows in a parking lot topology. Figure Figure 7 below shows the evolution of rate of these three flows. The figure shows the convergence of three flows with bottleneck bandwidth of 500kbps. Flow 2 starts 10 time unit later flow 1 and flow

3 falls behind 20 units from flow 1. The initial rates of these flows were set to be 20kbps, 250kbps and 500kbps respectively. We observe that max-min converges at 256kbps whereas proportional converges slower to 210kbps with much lesser loss rate.

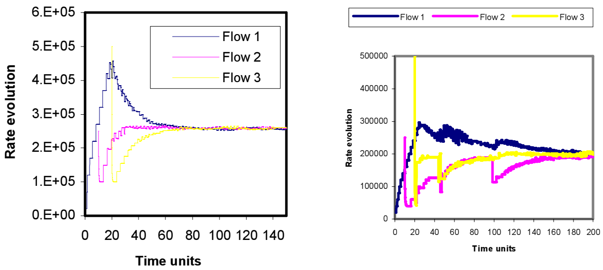

Figure 7 Rate evolution of max-min and proportional controllers

B. Proportional and Tangential AQM Feedback

Our tangential AQM feedback extends the proportional scheme with additive link penalty but imposes a well known penalty scaling factor to appropriately adjust the link losses. Our experiment setup and the control parameters remain the same as in earlier experiment. We observe several interesting features of our controller and the relative merits in the performance. Figure Figure 8 Rate evolution of proportional and tangential controllers

shows that the tangential controller is capable of achieving convergence to fairness much closer to the link bottleneck with much lesser packet loss. While proportional controller converges at 256kbps, our fair convergence occurs at around 200kbps for all three flows.

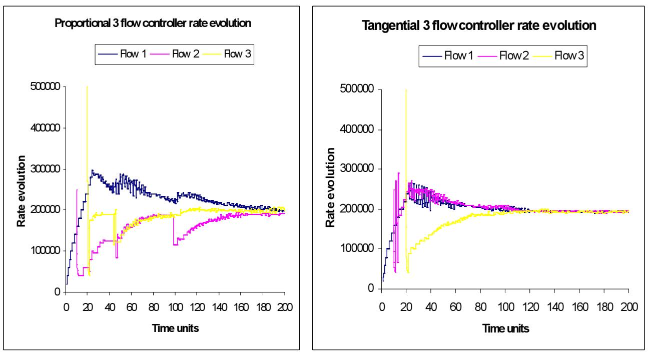

Figure 8 Rate evolution of proportional and tangential controllers

Figure 9 demonstrates the flow loss behavior across the two bottleneck links for proportional and tangential controllers. The aggregate positive scaling factor results in a much stable and smaller flow losses compared to the proportional controller.

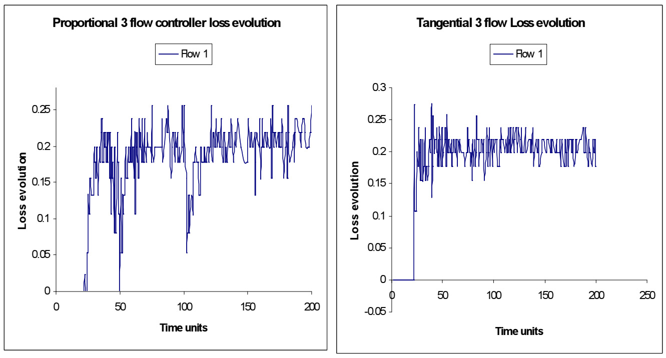

Figure 9 Flow loss evolution of proportional and tangential controllers

VII. CONCLUSIONS

In this paper, we considered a form of continuous feedback controller based on optimization theory that is proved to have several merits compared to the well-studied logarithmic proportional controller. We derive our motivation from sequential nonlinear programming method that allows additional penalty to objective function based on the square of linear constraint. Here, we showed that this additional penalty has immediate application to adjust our loss penalty function and network cost factor. Using this, we developed a novel tangential source rate controller whose trajectory followed closely that of source's own cost function and proved that this indeed minimizes the aggregate flow losses. Using simulations we also established its convergence and existence of stationary optimal rates that are more meaningful in parking-lot topologies. Finally, we establish the asymptotic stability with upper bounds on the increase and decrease parameters α, β, and β_c.

Recollecting some of the motivation in the earlier sections, we see that our scheme allows a well-defined correlation in rate adjustment. The significance of our work is its improvement in the speed of convergence and consistent reduction in packet loss up to 4 times lesser compared to logarithmic proportional controller for arbitrary topologies. The packet loss scaling factor is proved to be more meaningful and applicable to our network cost factor. Our tangential controller can thus be well suited for high bandwidth-delay product networks and large streaming applications where the optimum throughput usage is to be brought closer to the bottleneck bandwidth. In future, we intend to study the aggressiveness of this scaling factor and whether such penalty may be applicable for general form of utility functions.

VIII. REFERENCES

[1] S. Floyd, and K. Fall, "Promoting the Use of End-to-End Congestion Control in the Internet", *IEEE/ACM Transactions on Networking*, 1999.

[2] F. Kelly, A. Maulloo, and D. Tan, "Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability", *Journal of the Operational Research Society*, 1998.

[3] S. Kunniyur and R. Srikant, "End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks", *IEEE INFOCOM 2000*.

[4] L. Massoulié and J. Roberts, "Bandwidth Sharing: Objectives and Algorithms", *IEEE INFOCOM*, March 1999.

[5] T. Alpcan and T. Basar, "A Utility-Based Congestion Control Scheme for Internet-Style Networks with Delay", *IEEE INFOCOM*, San Francisco, CA, March 30-April 3, 2003.

[6] P. Marbach, "Priority service and max-min fairness", *IEEE INFOCOM*, 2002.

[7] A. Ganesh and K. Laevens, "Congestion Pricing and User Adaptation", *IEEE INFOCOM*, Anchorage, USA, April 2001.

[8] J. Padhye, V. Firoiu, D. Towsley and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation", *ACM SIGCOMM* 1998.

[9] S.H. Low and D. E. Lapsley, "Optimization flow control: Part I---Basic algorithms and convergence", *IEEE/ACM Transactions on Networking*, vol. 7, pp. 861-875, Dec. 1999.

[10] Rangarajan K. Sundaram. A First Course in Optimization Theory, Cambridge University Press, 1996.

[11] D. Katabi, M. Handley, and C. Rohrs, "Internet Congestion Control for High Bandwidth-Delay Product Networks", *ACM SIGCOMM*, 2002, Pittsburgh, August, 2002.

[12] C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, W. Feng, O. Martin, H. Newman, F. Paganini, S. Ravot, and S. Singh, "FAST TCP: From Theory to Experiments", *Submitted to IEEE Communications Magazine*, April, 2003.

[13] L. Tassiulas K. Kar, S. Sarkar, "A simple rate control algorithm for maximizing total user utility", *IEEE NFOCOM*, 2001.

[14] T. Alpcan and T. Basar. "A Utility-Based Congestion Control Scheme for Internet-Style Networks with Delay", *IEEE INFOCOM*, San Francisco, CA, March 30-April 3, 2003.

[15] J. Mo and J. Walrand, "Fair End-to-End Window-based Congestion Control". *IEEE/ACM Trans. on Networking* 8-5, pp. 556-567, 2000.

[16] R. J. La and V. Ananthram, "Charge-Sensitive TCP and Rate Control in the Internet", *IEEE INFOCOM* 2000, March 2000.

[17] Thomas L. Vincent, Walter J. Grantham, "Nonlinear and Optimal Control Systems", Wiley & Sons, Inc, July 1997.

[18] S. Ganesh and D. Loguinov, "Stability Analysis of Continuous Controllers: Heterogeneous Delays and Explicit Loss Dependence", Texas A&M Computer Science Technical Report, 2003.

[19] Robert M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, ``On the Lambert W Function", Advances in Computational Mathematics, volume 5, 1996.