Modelling and Stability of FAST TCP

Jiantao Wang   David X.Wei   Steven H. Low

California Institute of Technology, Pasadena, CA 91125, USA

{jiantao@cds, weixl@, slow@}caltech.edu

Abstract—We introduce a discrete-time model of FAST TCP that fully captures the effect of self-clocking, and compare it with the traditional continuous-time model. While the continuous-time model predicts instability for homogeneous sources sharing a single link when feedback delay is large, experiments suggest otherwise. Using the discrete-time model, we prove that FAST TCP is locally asymptotically stable in general networks when all sources have a common round-trip feedback delay, no matter how large the delay is. We also prove global stability for a single bottleneck link in the absence of feedback delay. The techniques developed here are new and applicable to other protocols.

Index Terms—Congestion control, FAST TCP, Local stability, Global stability

I. INTRODUCTION

Congestion control is a distributed feedback algorithm to allocate network resources among competing users. The algorithms in the current Internet, TCP Reno and its variants, have prevented severe congestion while the Internet underwent explosive growth during the last decade. However, it is well-known that these algorithms do not scale as bandwidth-delay product of the Internet continues to grow [8], [13]. This has motivated several recent proposals for congestion control of high speed networks, including HSTCP [6], STCP [11], FAST TCP [9], and BIC TCP [19] (see [9] for extensive references). Local stability of FAST TCP in the absence of feedback delay is proved in [9] for the case of a single link. In this paper, we extend the analysis to local stability for general networks with feedback delay and global stability on a single link without feedback delay.

Most of the stability analysis in the literature is based on the fluid model introduced in [8] (see surveys in [12], [10], [16] for extensions and related models). A key feature of many of these models is that a source controls its sending rate directly[1], and that the queueing delay at a link is proportional to the integral of the

[1]Even when the congestion window size is used as the control variable, sending rate is often taken to be the window normalized by a *constant* round-trip time, and hence a source still controls its rate directly.

excess demand for its bandwidth. In reality, a source dynamically sets its congestion window rather than its sending rate. These models do not adequately capture the self-clocking effect where a packet is sent only when an old one is acknowledged, except briefly and immediately after the congestion window is changed. This automatically constrains the input *rate* at a link to the link capacity, after a brief transient, no matter how large the congestion windows are set. Recently, a new discrete-time link model is proposed in [18], [9] to capture this effect, and experimental validations have been carried out in [18]. While the traditional continuous-time link model does not consider the self-clocking, our new discrete-time link model ignores the fast dynamics at the links. We first present both models in Section II. We then provide experimental results that show that, despite errors in these models, both of them track queueing delays reasonably well.

In Section III we compare the predictions on stability using the continuous-time model to experiments on the Dummynet testbed. When sources are homogeneous, i.e., have the same round trip delay, experiments so far suggest that FAST TCP is always stable on a single link, while the continuous-time model predicts instability when the delay is large.

We analyze the stability of FAST TCP using the discrete-time model in Sections IV and V. In Section IV, we prove the local asymptotic stability of FAST TCP in arbitrary networks when all sources have the *same* round-trip feedback delay, no matter how large the delay is. In Section V, we restrict ourselves to a single link without feedback delay and prove the global stability of FAST TCP. The techniques developed in this paper are new and applicable to analyzing other protocols.

Finally, we conclude in Section VI with limitations of this paper.

II. MODEL

A. Notation

A network consists of a set of L links indexed by l with finite capacity c_l. It is shared by a set of N unicast flows identified by their sources indexed by i. Let R be

the routing matrix where $R_{li} = 1$ if source i uses link l, and 0 otherwise.

FAST TCP updates its congestion window every fixed time period. We use this update period as time unit. All the time variables are measured in this time unit.

Let d_i denote the round-trip propagation delay of source i, and $q_i(t)$ denote the round-trip queueing delay. The round-trip time is given by $T_i(t) := d_i + q_i(t)$. We denote the forward feedback delay from source i to link l by τ_{li}^f, and the backward feedback delay from link l to source i as τ_{li}^b. The sum of forward delay from source i to any link l and the backward delay from link l to source i is fixed, i.e., $\tau_i := \tau_{li}^f + \tau_{li}^b$ for any link l on the path of source i. We make a subtle assumption here. In reality, the feedback delays τ_{li}^f, τ_{li}^b include queueing delay and are time-varying. We assume for simplicity that they are constants, and mathematically unrelated to $T_i(t)$. Later, when we analyze linear stability around the network equilibrium in the presence of feedback delay, we can interpret τ_i as the equilibrium value of T_i.

Let $w_i(t)$ be source i's congestion window at time t. The sending rate of source i at time t is defined as

$$x_i(t) = \frac{w_i(t)}{d_i + q_i(t)} \qquad (1)$$

The aggregate rate at link l is

$$y_l(t) = \sum_i R_{li} x_i(t - \tau_{li}^f) \qquad (2)$$

Let $p_l(t)$ be the queueing delay at link l. The end-to-end queueing delay $q_i(t)$ observed by source i is

$$q_i(t) = \sum_l R_{li} p_l(t - \tau_{li}^b) \qquad (3)$$

B. Discrete and continuous-time models

FAST TCP source periodically updates its congestion window w based on the average RTT and estimated queueing delay. The pseudo-code is:

$$w \leftarrow (1 - \gamma)w + \gamma \left(\frac{\text{baseRTT}}{\text{RTT}} w + \alpha \right)$$

where $\gamma \in (0, 1]$, baseRTT is the minimum RTT observed, and α is a constant.

We model this by the following discrete time equation

$$w_i(t+1) = \gamma \left(\frac{d_i w_i(t)}{d_i + q_i(t)} + \alpha_i \right) + (1 - \gamma) w_i(t) \quad (4)$$

where $w_i(t)$ is the congestion window of the ith source, $\gamma \in (0, 1]$, and α_i is a constant for source i. The corresponding continuous-time model is

$$\dot{w}_i(t) = \gamma \left(\frac{d_i w_i(t)}{d_i + q_i(t)} + \alpha_i - w_i(t) \right) \qquad (5)$$

where the time is measured in the unit of update period in FAST TCP.

For the continuous-time model, queueing delay has been traditionally modelled with

$$\dot{p}_l(t) = \frac{1}{c_l} (y_l(t) - c_l) \qquad (6)$$

However, TCP uses self-clocking: the source always tries to maintain that the number of packets in fly equals to the congestion window size. When the congestion window is fixed, the source will send a new packet exactly after it receives an ACK packet. When the congestion window changes, the source sends out bulk traffic in burst, or sends nothing in a short time period. Therefore, one round-trip time after a congestion window is changed, packet transmission will be clocked at the same rate as the throughput the flow receives. We assume that the disturbance in the queues due to congestion window changes settles down quickly compared with the update period of the discrete-time model; see [18] for detailed justification and validation experiments for these arguments. A consequence of this assumption is that the link queueing delay vector, $p(t) = (p_l(t)$, for all $l)$, is determined implicitly by sources' congestion windows in a static manner:

$$\sum_i R_{li} \frac{w_i(t - \tau_{li}^f)}{d_i + q_i(t - \tau_{li}^f)} \begin{cases} = c_l & \text{if } p_l(t) > 0 \\ \leq c_l & \text{if } p_l(t) = 0 \end{cases} \quad (7)$$

where the q_i is the end-to-end queueing delay given by (3).

In summary, the continuous-time model is specified by (5) and (6), and discrete-time model by (4) and (7), where the source rates and aggregate rates at links are given by (1) and (2), and the end-to-end delays are given by (3). While the continuous-time model does not take self-clocking into a full account, the discrete-time model ignores the fast dynamics at the links. Before comparing these models, we clarify their common equilibrium structure.

Theorem 1 ([9]). *Suppose that the routing matrix R has full row rank. A unique equilibrium (x^*, p^*) of the network exists, and x^* is the unique maximizer of*

$$\max_{x \geq 0} \sum_i \alpha_i \log x_i \quad s.t. \quad Rx \leq c \qquad (8)$$

with p^* as the corresponding optimum of its Lagrangian dual. This implies in particular that the equilibrium rate x^* is α_i-weighted proportionally fair.

C. Validation

The continuous-time link model implies that the queue takes infinitely amount of time to converge after a

window change. On the other extreme, the discrete-time link model assumes that the queue settles down in one sampling time. Neither is perfect, but we now present experimental results that suggest both track the queue dynamics well.

All the experiments reported in this paper are carried out on the Dummynet testbed [4]. A FreeBSD machine is configured as a Dummynet router which provides different propagation delays for different sources. It can be configured with different capacity and buffer size. In our experiments, the bottleneck link capacity is 800Mbps, and the buffer size is 4000 packets with fixed packet length of 1500 bytes. A Dummynet monitor records the queue size every 0.4 second. The congestion window size and RTT are recorded at the host every 50ms. TCP traffic is generated using *iperf*. The publicly released code of FAST [5] is used in all experiments involving FAST. We present two experiments to validate the model, one closed-loop and one open-loop.

In the first (closed-loop) experiment, there are 3 FAST TCP sources sharing a Dummynet router with a common propagation delay of 100ms. The measured, and predicted queue sizes are given in Figure 1. In the beginning of the experiment, the FAST sources are in the slow start phase, none of the models gives accurate prediction. After the FAST TCP enter the congestion avoidance phase, both models track the queue size well.
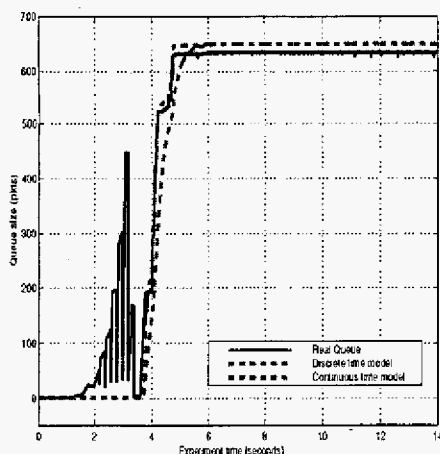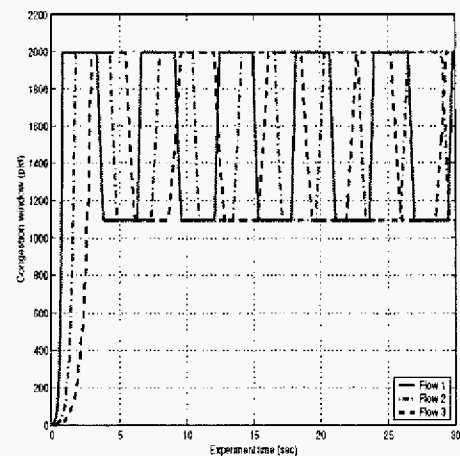

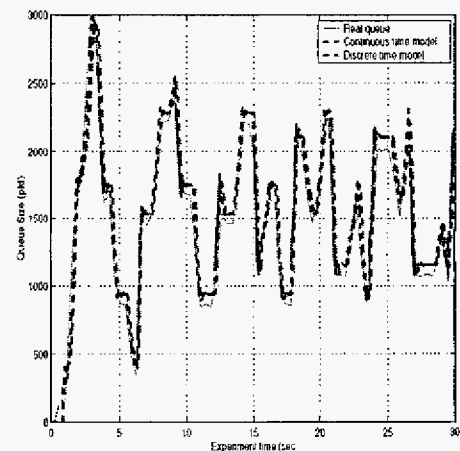Fig. 1.   Model validation: closed loop with FAST TCP

To eliminate the modelling error in congestion window adjustment algorithm itself while validating the link models, we decouple the TCP and queue dynamics by using open loop window control. The second experiment involves three sources with propagation delays 50ms, 100ms and 150ms sharing the same Dummynet router.

We changed the Linux 2.4.19 kernel so that the sources vary their window sizes according to the schedules shown in Figure 2(a). The sequences of congestion

window sizes are then used in (1)–(2) and (6) to compute the queueing delay predicted by the continuous-time model. We also use them in (1)–(2) and (7) to compute the predictions of the discrete-time model. The queueing delay measured from the Dummynet, and those predicted by these two models are shown in Figure 2(b) which indicates that both models track the queue sizes well. We next analyze the stability properties of these two models.

(a) Congestion windows

(b) Queue size

Fig. 2.   Model validation: congestion window are set according to the schedules in (a). Resulting queueing delays are shown in (b).

III. NUMERICAL SIMULATION AND EXPERIMENT WITH CONTINUOUS-TIME MODEL

Using the standard technique, we can derive a sufficient condition for local stability of FAST TCP for a general network using the continuous-time model. The condition implies that FAST TCP will become unstable when the ratio between the propagation delay and queueing delay is large. However our experiments of FAST TCP on the Dummynet testbed has always

been stable in the case of homogeneous sources and a single link even when the delay well exceeds the the threshold in the stability condition. We now present such an experiment which not only violates the local stability condition, moreover, numerical simulation of the continuous-time model exhibits instability. Yet, the same network on Dummynet is clearly stable. This suggests that the discrepancy is not in the stability theorem but rather in the continuous-time model.

In our experiment, the sources have identical propagation delay of 100ms with a constant α value of 70 packets. They share a bottleneck with capacity of 800Mbps. The simulations and experiments consist of three intervals. The interval length is 10 seconds for the continuous-time model simulation, and 100 seconds for the experiment[2]. Three sources are active from the beginning of the experiment, 7 additional sources activate in the second interval, and in the last interval, all sources become inactive except 5 of them. The simulation and experimental results are shown in Figure 3 and Figure 4 respectively.

When the ratio between the propagation delay and queueing delay is large, the numerical simulation based on the continuous-time model exhibits periodic oscillation. However, in the Dummynet experiment, FAST TCP is actually stable (see Figure 4).[3]

We believe that the discrepancy is largely due to the fact that the continuous-time model does not capture the self-clocking effect accurately. Self-clocking ensures that packets are sent at the same rate as the goodput the source receives, except briefly when the window size changes. This self-clocking feature can actually help the system approach to an equilibrium. Indeed, for the case of one source one link, a discrete-event model is used in [18] to prove that TCP FAST and Vegas is always stable regardless of the feedback delay. It also provides justification for the the discrete-time model (4) and (7) based on the self-clocking feature introduced in the last section. We now analyze the stability of this model. We will see that the discrete-time model predicts that a network of homogeneous sources with the same feedback delay is locally stable no matter how large the delay is, agreeing with our experimental experience.
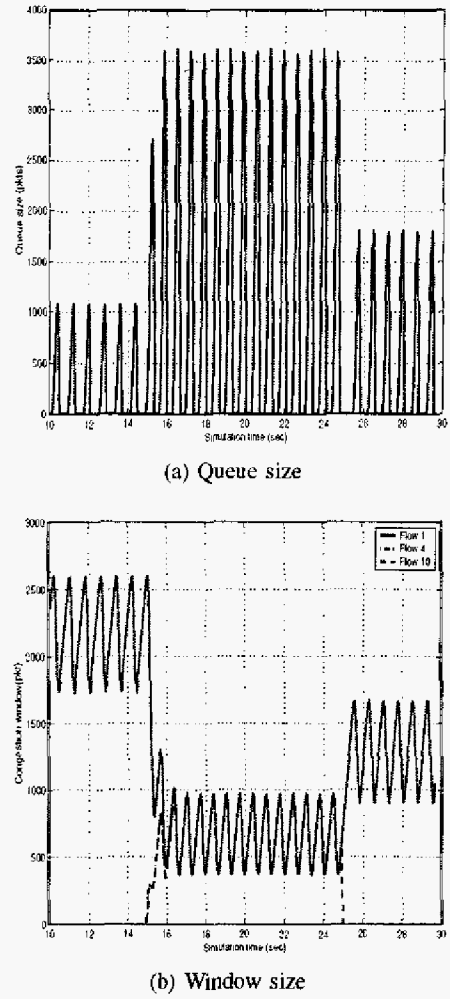

(a) Queue size

(b) Window size

Fig. 3. Numerical simulations of FAST TCP

IV. LOCAL STABILITY IN ARBITRARY NETWORK WITH COMMON FEEDBACK DELAY

A network of FAST TCP sources is modelled by equations (3), (4), (7). This generalizes the model in [9] by including feedback delay. When local stability is studied, we ignore all un-congested links (links where prices are zero in equilibrium) and assume that equality always holds in (7).

The main result of this section provides a sufficient condition for local asymptotic stability in general networks with common feedback delay.

Theorem 2. *FAST TCP is locally stable for arbitrary networks if* $\gamma \in (0,1]$ *and all sources have the same round-trip feedback delay* $\tau_i = \tau$ *for all* i.

The stability condition in the theorem does not depend on the value of the feedback delay, but only on the heterogeneity among them. In particular, when all feedback delays are ignored, $\tau_i = 0$ for all i, then FAST TCP
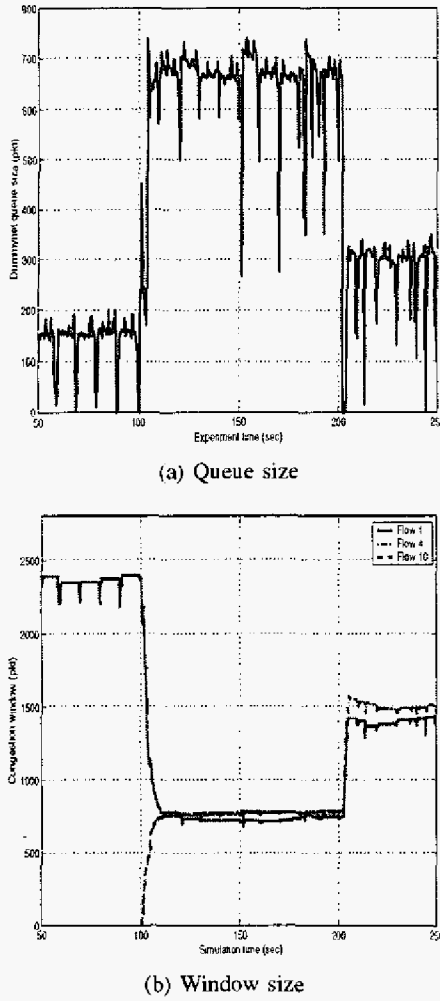
---

[2]We use a long duration in the Dummynet experiment because a FAST TCP source takes longer to converge due to slow-start which is not included in our model.

[3]The regular spikes every 10 seconds in the queue size is probably due to certain background task in the sending host.

The relation $\tau_{li}^f + \tau_{li}^b = \tau_i$ gives

$$R_b(z) = R_f(z^{-1}) \cdot \text{diag}(z^{-\tau_i}) \qquad (9)$$

Denote $W(z)$, $Q(z)$ and $P(z)$ as the corresponding Z-transforms of $\delta w(t)$, $\delta q(t)$ and $\delta p(t)$ for the linearized system. Let q and w be the end-to-end queueing delay and congestion window at equilibrium. Linearizing (7) yields

$$\sum_i R_{li} \left(\frac{\delta w_i(t - \tau_{li}^f)}{d_i + q_i} - w_i \frac{\delta q_i(t - \tau_{li}^f)}{(d_i + q_i)^2} \right) = 0$$

where the equality is used in (7). The corresponding Z-transform in matrix form is

$$R_f(z)D^{-1}MW(z) - R_f(z)BQ(z) = 0 \qquad (10)$$

where the diagonal matrices B, D and M are:

$$B := \text{diag}\left(\frac{w_i}{(d_i + q_i)^2} \right) , M := \text{diag}\left(\frac{d_i}{d_i + q_i} \right)$$
$$D := \text{diag}(d_i)$$

Since $R_f(z)$ is generally not a square matrix, we cannot cancel it in (10).

Equation (3) is already linear, and the corresponding Z-transform in matrix form is

$$Q(z) = R_b(z)^T P(z) \qquad (11)$$

By combining (10) and (11), we obtain

$$\begin{pmatrix} I & -R_b^T(z) \\ R_f(z)B & 0 \end{pmatrix} \begin{pmatrix} Q(z) \\ P(z) \end{pmatrix}$$
$$= \begin{pmatrix} 0 \\ R_f(z)D^{-1}M \end{pmatrix} W(z)$$

Solving this equation with block matrix inverse gives the transfer function from $W(z)$ to $Q(z)$:

$$\frac{Q(z)}{W(z)} = R_b^T(z)(R_f(z)BR_b^T(z))^{-1}R_f(z)D^{-1}M$$

The Z-transform of the linearized congestion window update algorithm of FAST TCP is

$$zW(z) = \gamma(MW(z) - DBQ(z)) + (1 - \gamma)W(z)$$

By combining the above equations, we get the open loop transfer function $L(z)$ from $W(z)$ to $W(z)$ as

$$L(z) = -\left(\gamma \left(M - DBR_b^T(z)(R_f(z)BR_b^T(z))^{-1} \right. \right.$$
$$\left. \left. R_f(z)D^{-1}M \right) + (1 - \gamma)I \right) z^{-1}$$

A sufficient condition for local stability can be developed for FAST TCP based on the generalized Nyquist criterion [3], [2]. Since any changes in the input w can at most give a finite change at the output. Therefore, the system is open loop stable. If we can show that the eigenvalue

(a) Queue size

(b) Window size

Fig. 4. Dummynet experiments of FAST TCP

is locally asymptotically. This generalizes the stability result in [9].

Corollary 3. *FAST TCP is locally asymptotically stable in the absence of feedback delay for general networks with any* $\gamma \in [0, 1)$.

The rest of this section is devoted to the proof of Theorem 2.

We apply Z-transform to the linearized system, and use the generalized Nyquist criterion to derive a sufficient stability condition. Define the forward and backward Z-transformed routing matrices $R_f(z)$ and $R_b(z)$ as:

$$(R_f(z))_{li} := \begin{cases} z^{-\tau_{li}^f} & \text{if } R_{li} = 1 \\ 0 & \text{if } R_{li} = 0 \end{cases}$$

$$(R_b(z))_{li} := \begin{cases} z^{-\tau_{li}^b} & \text{if } R_{li} = 1 \\ 0 & \text{if } R_{li} = 0 \end{cases}$$

loci of $L(e^{jw})$ does not enclose -1 for $\omega \in [0, 2\pi)$, the closed loop system is stable. Therefore, if the spectral radius of $L(e^{jw})$ is strictly less than 1 for $\omega \in [0, 2\pi)$, the system will be stable.

When $z = e^{jw}$, the spectral radii of $L(z)$ and $-zL(z)$ are the same. Hence, we only need to study the spectral radius of

$$J(z) : = \gamma(M - DBR_b^T(z)\left(R_f(z)BR_b^T(z)\right)^{-1}$$
$$R_f(z)D^{-1}M + (1 - \gamma)I$$

Clearly, the eigenvalues of $J(z)$ are dependent on γ. For any given $z = e^{j\omega}$, let the eigenvalues of $J(z)$ be denoted by $\lambda_i(\gamma)$, $i = 1 \ldots N$, as functions of $\gamma \in (0, 1]$. It is clear that:

$$|\lambda_i(\gamma)| = |\gamma\lambda_i(1) + (1 - \gamma)|$$
$$\leq \gamma|\lambda_i(1)| + (1 - \gamma)$$

Hence if $\rho(J(z)) < 1$ for any $z = e^{j\omega}$ for $\gamma = 1$, it will also hold for all $\gamma \in (0, 1]$. Therefore, it suffices to study stability condition for $\gamma = 1$.

Let μ_i be the ith diagonal entry of matrix M with $\mu_i = d_i/(d_i + q_i)$. Denote $\mu_{\max} := \max_i \mu_i$. Since the end-to-end queueing delay q_i cannot be zero at equilibrium (otherwise the rate will be infinitely large), we have $q_i > 0$ and $\mu_{\max} < 1$. The following lemma characterizes the eigenvalues of $J(z)$ with $\gamma = 1$.

Lemma 4. *When $z = e^{j\omega}$ with $\omega \in [0, 2\pi)$ and $\gamma = 1$, the eigenvalues of $J(z)$ have the following properties:*

1) *There are L zero eigenvalues with the corresponding eigenvectors as the columns of matrix $M^{-1}DBR_b^T(z)$.*
2) *The nonzero eigenvalues have moduli less than 1 if $\tau_{\max} - \tau_{\min} < 1/4$, where $\tau_{\max} = \max_i \tau_i$ and $\tau_{\min} = \min_i \tau_i$*

Proof: At $\gamma = 1$, the matrix $J(z)$ is

$$M - DBR_b^T(z)(R_f(z)BR_b^T(z))^{-1}R_f(z)D^{-1}M$$

It is easy to check that:

$$J(z)M^{-1}DBR_b^T(z) = DBR_b^T(z) - DBR_b^T(z)$$
$$= 0$$

Since $M^{-1}DBR_b^T(z)$ has full column rank, it consists of L linearly independent eigenvectors of $J(z)$ with corresponding eigenvalue 0. This proves the first assertion.

For the second assertion, suppose that λ is an eigenvalue of $J(z)$ for a given z. Define matrix A as

$$A : = J(z) - \lambda I$$
$$= (M - \lambda I) - DBR_b^T(z)(R_f(z)BR_b^T(z))^{-1}$$
$$R_f(z)D^{-1}M$$

which is singular by definition. From the matrix inversion formula (see e.g. [7])

$$(J + EHS)^{-1}$$
$$= J^{-1} - J^{-1}E(H^{-1} + SJ^{-1}E)^{-1}SJ^{-1}$$

if $J + EHS$ is singular, then either J or $H^{-1} + SJ^{-1}E$ is singular. We can let

$$J := M - \lambda I, \qquad E := -DBR_b^T(z),$$
$$H := (R_f(z)BR_b^T(z))^{-1}, \qquad S := R_f(z)D^{-1}M.$$

Since $A = J + EHS$ is singular, either $J = M - \lambda I$ or $H^{-1} + SJ^{-1}E$ is singular. The second term can be reformulated into

$$R_f(z)(B - M(M - \lambda I)^{-1}B)R_b^T(z)$$

Case 1: $M - \lambda I$ is singular. Since M is diagonal, then

$$0 < \lambda = \frac{d_i}{d_i + q_i} = \mu_i \leq \mu_{\max} < 1$$

Case 2: $R_f(z)(B - M(M - \lambda I)^{-1}B)R_b^T(z)$ is singular. It is clear that

$$B - M(M - \lambda I)^{-1}B$$
$$= \text{diag}\left(1 - \mu_i(\mu_i - \lambda)^{-1}\beta_i\right)$$
$$= -\lambda\text{diag}\left(\frac{\beta_i}{\mu_i - \lambda}\right)$$

where β_i is the ith diagonal entry of matrix B. Hence, $\lambda = 0$ is always an eigenvalue, which is claimed before. If λ is nonzero, it has to be true that

$$\det\left(R_f(z)\text{diag}\left(\frac{\beta_i}{\mu_i - \lambda}\right)R_b^T(z)\right) = 0 \qquad (12)$$

When $z = e^{jw}$, we have $z^{-1} = \bar{z}$. Hence, equation (9) can be rewritten as

$$R_b^T(z) = \text{diag}(z^{-\tau_i})R_f^T(\bar{z})$$
$$= \text{diag}(z^{-\tau_i})R_f^*(z)$$

When $z = e^{jw}$, the nonzero eigenvalue λ satisfies

$$\det\left(R_f(z)\text{diag}\left(\frac{e^{j(\theta_i + \psi)}\beta_i}{\mu_i - \lambda}\right)R_f^*(z)\right) = 0$$

Substituting the above equation into (12) with $z = e^{j\omega}$ yields

$$\det\left(R_f(z)\text{diag}\left(\frac{e^{-j\omega\tau_i}\beta_i}{\mu_i - \lambda}\right)R_f^*(z)\right) = 0 \qquad (13)$$

Therefore, the following formula is also zero

$$e^{-j(\omega\tau_{\max} + \psi)}\det\left(R_f(z)\text{diag}\left(\frac{e^{j(\theta_i + \psi)}\beta_i}{\mu_i - \lambda}\right)R_f^*(z)\right)$$

where $\theta_i = (\tau_{\max} - \tau_i)\omega$, and ψ can be any value. When $\tau_{\max} - \tau_{\min} < 1/4$, we have

$$0 \le \theta_i = (\tau_{\max} - \tau_i)\omega \le \pi/2$$

Suppose that there is a solution such that $|\lambda| \ge 1$. From Lemma 5 below, there exists a ψ s.t. $\mathrm{Im}(\mathrm{diag}\left(e^{j(\theta_i+\psi)}\beta_i/(\mu_i - \lambda)\right))$ is a positive diagonal matrix. Therefore the imaginary part of matrix $R_f(z)\mathrm{diag}\left(e^{j(\theta_i+\psi)}\beta_i/(\mu_i - \lambda)\right))R_f^*(z)$ is positive definite, and the real part is symmetric. From Lemma 6 below, it has to be nonsingular. This contradicts the equation

$$\det\left(R_f(z)\mathrm{diag}\left(\frac{e^{j(\theta_i+\psi)}\beta_i}{\mu_i - \lambda}\right)R_f^*(z)\right) = 0$$

Hence, we have $|\lambda| < 1$. □

The proof of Theorem 2 will be complete after the next two Lemmas.

Lemma 5. *Suppose that $0 < \mu_i < 1$ and $0 \le \theta_i < \pi/2$. If $|\lambda| \ge 1$, there exists a ψ such that*

$$Im\left(\frac{e^{j(\theta_i+\psi)}\beta_i}{\mu_i - \lambda}\right) > 0 \quad for \quad i = 1 \dots N$$

Proof: See Appendix VII-A. □

Lemma 6. *If the real part of a complex matrix is symmetric, and the imaginary part is positive definite, then the matrix is nonsingular.*

Proof: See Appendix VII-B. □

V. GLOBAL STABILITY FOR ONE LINK WITHOUT FEEDBACK DELAY

In the absence of feedback delay, when there is only one link, the FAST TCP model can be simplified into

$$w_i(t+1) = \gamma\left(\frac{d_i w_i(t)}{d_i + q(t)} + \alpha_i\right) + (1-\gamma)w_i(t) \quad (14)$$

$$\sum_i \frac{w_i(t)}{d_i + q(t)} \le c \quad \text{with equality if } q(t) = 0 \quad (15)$$

where $q(t)$ is the queueing delay at the link (subscript is omitted). The main result of this section proves that the above system (14)–(15) is globally asymptotically stable and converges to the equilibrium exponentially fast starting from any initial value.

Theorem 7. *On a single link, FAST TCP converges exponentially to the equilibrium, in the absence of feedback delay.*

In the rest of this section, we prove the theorem in several steps. The first result is that after finite steps K_1, equality always holds in (15) and $q(t) > 0$ for any $t > K_1$. Define the normalized congestion window sum as

$$Y(t) := \sum_i \frac{w_i(t)}{d_i} \quad (16)$$

From (15), it is clear that $q(t) > 0$ if and only if $Y(t) > c$.

Lemma 8. *There exists $K_1 > 0$ such that the following claims are true for all $t > K_1$:*

1) $q(t) > 0$
2) $\nu(t+1) = (1-\gamma)\nu(t)$ *where*

$$\nu(t) := Y(t) - c - \sum_i \frac{\alpha_i}{d_i} \quad (17)$$

Proof: If initially $q(t) = 0$ which also means $Y(t) \le c$, from (14) we have

$$Y(t+1) = Y(t) + \gamma\sum_i \frac{\alpha_i}{d_i} \quad (18)$$

which linearly increases with t. Then $Y(t) > c$ after some finite steps, and there is a K_1 such that $Y(t) > c$ and $q(t) > 0$ at $t = K_1$

We now prove that $Y(t) > c$ implies $Y(t+1) > c$. Hence $q(t) > 0$ for $t > K_1$. Moreover, $\nu(t)$ converges exponentially to 0.

Suppose $Y(t) > c$. From $\sum_i w_i(t)/(d_i + q_i(t)) = c$, we have

$$
\begin{aligned}
\nu(t+1) &= \sum_i \frac{w_i(t+1)}{d_i} - \sum_i \frac{\alpha_i}{d_i} - c \\
&= (1-\gamma)\sum_i \frac{w_i(t) - \alpha_i}{d_i} + \gamma\sum_i \frac{w_i(t)}{d_i + q(t)} - c \\
&= (1-\gamma)\left(\sum_i \frac{w_i}{d_i} - c - \sum_i \frac{\alpha_i}{d_i}\right) \\
&= (1-\gamma)\,\nu(t)
\end{aligned}
$$

This proves the second assertion. Moreover it implies

$$Y(t+1) = (1-\gamma)Y(t) + \gamma\left(\sum_i \frac{\alpha_i}{d_i} + c\right) \quad (19)$$

Hence, $Y(t) > c$ implies $Y(t+1) > c$ and $q(t+1) > 0$. This completes the proof. □

For the rest of this subsection, we pick a fixed ϵ with $0 < \epsilon < \sum_i \alpha_i/d_i$. Define

$$q_{\min} := \frac{d_{\min}}{c}\left(\sum_i \frac{\alpha_i}{d_i} - \epsilon\right)$$

$$q_{\max} := \frac{d_{\max}}{c}\left(\sum_i \frac{\alpha_i}{d_i} + \epsilon\right)$$

where $d_{\min} := \min_i d_i$ and $d_{\max} := \max_i d_i$. Then $q(t)$ is bounded by these two values after finite steps.

Lemma 9. *There exists a positive K_2 such that for any $t \geq K_2$ we have*

$$q_{\min} \leq q(t) \leq q_{\max}$$

Proof: From Lemma 8, after finite steps K_1, $\nu(t+1) = (1 - \gamma)\nu(t)$. Therefore, there exists a K_2 such that $|\nu(t)| < \epsilon$ for all $t \geq K_2$. It implies

$$\sum_i \frac{\alpha_i}{d_i} < \sum_i \frac{w_i(t)}{d_i} - c + \epsilon$$

$$= \sum_i \left(\frac{w_i(t)}{d_i} - \frac{w_i(t)}{d_i + q(t)}\right) + \epsilon$$

$$\leq \sum_i \frac{q(t)w_i(t)}{d_{\min}(d_i + q(t))} + \epsilon$$

$$= \frac{q(t)c}{d_{\min}} + \epsilon$$

Hence,

$$q(t) \geq \frac{d_{\min}}{c}\left(\sum_i \frac{\alpha_i}{d_i} - \epsilon\right) = q_{\min} \qquad (20)$$

The proof for $q_{\max}$ is the same. $\square$

Define $\mu_i(t) := d_i/(d_i + q(t))$, and denote $\mu_{\max} := \max_i d_i/(d_i+q_{\min})$, $\mu_{\min} := \min_i d_i/(d_i+q_{\max})$. Based on Lemma 9, it is true that

$$1 > \mu_{\max} \geq \mu_i(t) > \mu_{\min} > 0 \quad \text{for any } t \geq K_2$$

Define

$$\eta_i(t) := \frac{w_i(t) - \alpha_i}{\alpha_i d_i} - \frac{1}{q(t)}, \qquad (21)$$

and denote $\eta_{\max}(t) := \max_i \eta_i(t)$, $\eta_{\min}(t) := \min_i \eta_i(t)$. We will show that the congestion window update for source i is proportional to $\eta_i(t)$, and the system is at equilibrium if and only if all $\eta_i(t)$ are zero. The next Lemma gives bounds on $\eta_i(t)$.

Lemma 10. *There are two positive numbers δ_1 and δ_2 such that for all $t \geq K_2$*

$$\eta_{\max}(t) > -\delta_1(1 - \gamma)^t \qquad \eta_{\min}(t) < \delta_2(1 - \gamma)^t$$

Proof: From (19), it is easy to check that

$$Y(t+1) - Y(t) = -\gamma\nu(t)$$

By Lemma 8, when $t \geq K_2$ we have

$$Y(t+1) - Y(t) = -\gamma\nu(t)$$
$$\leq \gamma(1 - \gamma)^{t-K_2}|\nu(K_2)|$$
$$= \kappa(1 - \gamma)^t \qquad (22)$$

where $\kappa := \gamma(1 - \gamma)^{-K_2}|\nu(K_2)|$

The update of source i's congestion window is:

$$w_i(t+1) - w_i(t)$$
$$= \gamma\left(\frac{d_i w_i(t)}{d_i + q(t)} + \alpha_i - w_i(t)\right)$$
$$= \frac{\gamma q(t)}{d_i + q(t)}\left(\frac{\alpha_i d_i}{q(t)} - (w_i(t) - \alpha_i)\right)$$
$$= -\gamma\alpha_i d_i \frac{q(t)}{d_i + q(t)}\left(\frac{w_i(t) - \alpha_i}{\alpha_i d_i} - \frac{1}{q(t)}\right)$$
$$= -\gamma\alpha_i q(t)\mu_i(t)\eta_i(t)$$

Choose δ_1 large enough such that

$$\delta_1 N\gamma\alpha_{\min}q_{\min}\mu_{\min}/d_{\max} > \kappa$$

where $\alpha_{\min} := \min_i \alpha_i$. We now prove $\eta_{\max}(t) > -\delta_1(1-\gamma)^t$ for all $t \geq K_2$ by contradiction. Suppose that there is a time $t \geq K_2$ such that $\eta_{\max}(t) \leq -\delta_1(1-\gamma)^t$. Then all the $\eta_i(t)$ are negative, which implies

$$Y(t+1) - Y(t)$$
$$= \sum_i (w_i(t+1) - w_i(t))/d_i$$
$$= \sum_i -\gamma\alpha_i q(t)\mu_i(t)\eta_i(t)/d_i$$
$$\geq N(-\eta_{\max})\gamma\alpha_{\min}q_{\min}\mu_{\min}/d_{\max}$$
$$\geq \delta_1 N(1 - \gamma)^t\gamma\alpha_{\min}q_{\min}\mu_{\min}/d_{\max}$$
$$> \kappa(1 - \gamma)^t$$

This contradicts equation (22), which proves the claim. The proof for $\eta_{\min}(t)$ is similar. $\square$

Define $L(t)$ as:

$$L(t) := \eta_{\max}(t) - \eta_{\min}(t) \qquad (23)$$

The following Lemma implies that the difference between different $\eta_i(t)$ goes to zero exponentially fast.

Lemma 11. *There are two positive numbers δ_3 and δ_4, such that for $t \geq K_2$ we have*

1) $L(t) \geq 0$
2) $L(t+1) \leq (1 - \gamma + \gamma\mu_{\max})L(t) + \delta_3(1 - \gamma)^t$,
3) $L(t) \leq \delta_4(1 - \gamma + \gamma\mu_{\max})^t$

Proof: See Appendix VII-C. □

Lemma 12. *Both* $\eta_{\max}(t)$ *and* $\eta_{\min}(t)$ *converge to zero exponentially.*

Proof: When $t \geq K_2$, combining Lemma 10 and Lemma 11 yields an upper bound for $\eta_{\max}(t)$,

$$\begin{aligned}\eta_{\max}(t) &= L(t) + \eta_{min}(t) \\ &\leq \delta_4(1 - \gamma + \gamma\mu_{\max})^t + \delta_2(1 - \gamma)^t\end{aligned}$$

The lower bound of $\eta_{\max}$ is

$$-\delta_1(1 - \gamma)^t \leq \eta_{max}(t)$$

Since both the upper and lower bounds of $\eta_{\max}(t)$ converge to zero exponentially fast, it exponentially goes to zero. The proof for $\eta_{\min}(t)$ is similar. □

Proof of Theorem 7: The system is at equilibrium if and only if $w_i(t) = w_i(t+1)$ for all i. This is equivalent to $\eta_i(t) = 0$ for all i because of the equation

$$w_i(t + 1) - w_i(t) = -\gamma\alpha_i q(t)\mu_i(t)\eta_i(t)$$

Since both $\eta_{\max}(t)$ and $\eta_{\min}(t)$ converge to zero exponentially from any initial value, the system converges to the equilibrium defined by $\eta_i(t) = 0$ globally. □

VI. CONCLUSION

We have introduced a new discrete-time link model that fully captures the effect of self-clocking and compared it with the traditional continuous-time model. Using this model, we have derived a sufficient condition for local asymptotic stability for general networks in the presence of feedback delay. The condition states that the system is stable if the difference among delays of the sources is small. This implies, in particular, that a network with homogeneous sources is always stable, consistent with our experimental experience so far. We have proved that FAST TCP is globally stable on a single link in the absence of feedback delay.

This work can be extended in several ways. First, the condition for local asymptotic stability derived appears more restrictive than our experiments suggest. Moreover, we have also found scenarios where predictions of the discrete-time model disagree with experiment. This discrepancies should be clarified. Second, it will be interesting to extend the global stability analysis to general networks and in the presence of feedback delay. Finally, the new model and the analysis techniques here can be applied to analyze other congestion control algorithms.

REFERENCES

[1] Hyojeong Choe and Steven H. Low. Stabilized Vegas. In *Proc. of IEEE Infocom*, April 2003. http://netlab.caltech.edu.

[2] Desoer, C.A. and Y.T.Yang On the generalized Nyquist stability criterion IEEE Transaction on Automatic Control 25, 187-196

[3] Frank M.Calllier and Charles A. Desoer Linear System Theory 368-374 Springer-Verlag, New York. 1991

[4] Luigi Rizzo. Dummynet. http://http://info.iet.unipi.it/~luigi/ip_dummynet/.

[5] FAST Team FAST implementation code http://netlab.caltech.edu/FAST/

[6] Sally Floyd HighSpeed TCP for large congestion windows Internet draft draft-floyd-tcp-highspeed-02.txt work in progress.http://www.icir.org/floyd/hstcp.html

[7] Roger A.Horn and Charles R.Johnson Matrix Analysis p.18-p.19 Cambridge University Press 1985

[8] C.V.Hollot, V.Misra, D.Towsley, and W.B.Gong Analysis and design of controller for AQM routers supporting TCP flows IEEE Transaction on Automatic Control, 47(6) 2002.

[9] Cheng Jin, David X. Wei and Steven H. Low FAST TCP: motivation, architecture, algorithms, performance IEEE Infocom, March 2004, Hong Kong. http://netlab.caltech.edu

[10] Frank P. Kelly. Fairness and stability of end-to-end congestion control. *European Journal of Control*, 9:159–176, 2003.

[11] Tom Kelly Scalable TCP: Improving performance in high-speed wide area networks Submitted for publication http://www-lce.eng.cam.ac.uk/~ctk21/scalable/

[12] Steven H. Low, Fernando Paganini, and John C. Doyle. Internet congestion control. *IEEE Control Systems Magazine*, 22(1):28–43, February 2002.

[13] S. H. Low, F. Paganini, J. Wang, and J. C. Doyle. Linear stability of TCP/RED and a scalable control. *Computer Networks Journal*, 43(5):633–647, 2003. http://netlab.caltech.edu.

[14] Vishal Misra, Weibo Gong and Don Towsley Fluid-based Analysis of a Network of AQM Routers Supporting TCP Flows with an Application to RED SIGCOMM 2000

[15] Fernando Paganini, Zhikui Wang, John C. Doyle, and Steven H. Low. Congestion control for high performance, stability and fairness in general networks. *IEEE/ACM Transactions on Networking*, to appear in 2004. http://www.ee.ucla.edu/~paganini.

[16] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.

[17] Glenn Vinnicombe On the stability of networks operating TCP-like protocols In *Proc. of IFAC* 2002 http://netlab.caltech.edu/pub/papers/gv_ifac.pdf.

[18] David X. Wei. Congestion Control Algorithms for High Speed Long distance TCP Connections Master Thesis, Caltech, 2004. http://www.cs.caltech.edu/~weixl/research/msthesis.ps

[19] Lisong Xu, Khaled Harfoush, and Injong Rhee Binary Increase Congestion Control for Fast Long-Distance Networks IEEE Infocom, March 2004, Hong Kong.

VII. APPENDIX

A. Proof of Lemma 5

Proof: There is a complex plane in Figure 5. Let the points A, B, and λ represent the value of $\mu_{\min}$, $\mu_{\max}$ and λ, respectively. Z is the intersection of segment $A\lambda$ and
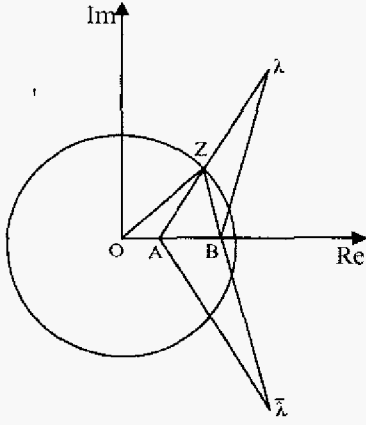
Fig. 5. Illustration of Lemma 5.

the unit circle, and $\overline{\lambda}$ stands for the complex conjugate of λ.

Let $\phi_i \in [0, 2\pi)$ be the phase of $1/(\mu_i - \lambda)$. Clearly, $\phi_i \in [0, \pi)$ if $\text{Im}(\lambda) \leq 0$, and $\phi_i \in (\pi, 2\pi)$ otherwise. Denote $\phi_{max} := \max_i \phi_i$ and $\phi_{min} := \min_i \phi_i$, then $0 \leq \phi_{max} - \phi_{min} \leq \pi$. Since every μ_i is in the range $[\mu_{min}, \mu_{max}]$, it is easy to check that every ϕ_i is in the range formed by the phases of $1/(\mu_{min} - \lambda)$ and $1/(\mu_{max} - \lambda)$. This implies

$$\phi_{max} - \phi_{min} \leq |\angle \frac{1}{\mu_{min} - \lambda} - \angle \frac{1}{\mu_{max} - \lambda}|$$
$$= \angle A\overline{\lambda}B = \angle A\lambda B$$
$$< \angle OZB < \pi/2$$

Let $\epsilon > 0$ be small enough such that $\phi_{max} - \phi_{min} < \pi/2 - \epsilon$. Choosing $\psi = -\phi_{min} + \epsilon$ gives us

$$\angle \frac{e^{j(\psi+\theta_i)}\beta_i}{\mu_i - \lambda} = \phi_i + \psi + \theta_i$$
$$= \phi_i - \phi_{min} + \epsilon + \theta_i \quad \text{(greater than 0)}$$
$$< \phi_{max} - \phi_{min} + \epsilon + \pi/2$$
$$< \pi$$

The fact that its phase is in $(0, \pi)$ implies that

$$Im\left(\frac{e^{j(\psi+\theta_i)}\beta_i}{\mu_i - \lambda}\right) > 0$$

$\square$

B. Proof of Lemma 6

Suppose that $A = A_r + jA_i$ where $A_r = A_r^T$ and A_i is positive definite. If A is singular, there exists a nonzero vector v such that $Av = 0$. Suppose that $v = \alpha + j\beta$. Then $Av = 0$ gives

$$A_r\alpha - A_i\beta = 0 \quad (24)$$
$$A_r\beta + A_i\alpha = 0 \quad (25)$$

Multiplying β^T to equation (24) yields

$$\beta^T A_r \alpha = \beta^T A_i \beta \geq 0 \quad (26)$$

Multiplying α^T to equation (25) gives us

$$\alpha^T A_r \beta = -\alpha^T A_i \alpha \leq 0 \quad (27)$$

Since $\beta^T A_r \alpha = \alpha^T A_r^T \beta = \alpha^T A_r \beta$, both (26) and (27) must hold with equality. This means that both α and β are zero. It contradicts the assumption that v is nonzero.

$\square$

C. Proof of Lemma 11

It is obvious that $L(t) \geq 0$ because of its definition in (23). We start with the update of $\eta_i(t)$ described as

$$\eta_i(t+1) - \eta_i(t)$$
$$= \frac{w_i(t+1) - w_i(t)}{\alpha_i d_i} - \frac{1}{q(t+1)} + \frac{1}{q(t)}$$
$$= -\frac{\gamma \alpha_i q(t) \mu_i(t) \eta_i(t)}{\alpha_i d_i} - \frac{1}{q(t+1)} + \frac{1}{q(t)}$$
$$= -\frac{\gamma q(t) \eta_i(t)}{d_i + q(t)} - \frac{1}{q(t+1)} + \frac{1}{q(t)}$$
$$= -\gamma(1 - \mu_i(t))\eta_i(t) - \frac{1}{q(t+1)} + \frac{1}{q(t)}$$

For simplicity, we let $a_i(t) := 1 - \gamma + \gamma\mu_i(t)$ and denote $a_{max} := 1 - \gamma + \gamma\mu_{max}$, then $a_i(t) \leq a_{max}$. This definition simplifies the above equation into

$$\eta_i(t+1) = a_i(t)\eta_i(t) - \frac{1}{q(t+1)} + \frac{1}{q(t)} \quad (28)$$

By comparing equation (28) for source i and j, we obtain

$$\eta_i(t+1) - \eta_j(t+1) = a_i(t)\eta_i(t) - a_j(t)\eta_j(t) \quad (29)$$

Without loss of generality, suppose that at time $t+1$, the largest and smallest values of η are achieved at sources i and j respectively. This assumption implies

$$L(t+1) = \eta_i(t+1) - \eta_j(t+1)$$

The upper bound of $L(t+1)$ is derived by considering separately the following three cases.

Case 1: $\eta_i(t)$ and $\eta_j(t)$ have different signs. It is easy to see that

$$L(t+1) = a_i(t)\eta_i(t) - a_j(t)\eta_j(t)$$
$$\leq a_{max}(\eta_i(t) - \eta_j(t))$$
$$\leq a_{max}(\eta_{max}(t) - \eta_{min}(t))$$
$$= a_{max}L(t)$$

Case 2: Both $\eta_i(t)$ and $\eta_j(t)$ are positive. It yields

$$
\begin{aligned}
L(t+1) &= a_i(t)\mu_i(t)\eta_i(t) - a_j(t)\eta_j(t) \\
&\leq a_{\max}\eta_{\max}(t) \\
&= a_{\max}L(t) + a_{\max}\eta_{\min}(t) \\
&\leq a_{\max}L(t) + a_{\max}\delta_2(1-\gamma)^t \\
&\leq a_{\max}L(t) + \delta_3(1-\gamma)^t
\end{aligned}
$$

where the last step is by choosing δ_3 larger than $a_{\max}\delta_2$.
Case 3: Both $\eta_i(t)$ and $\eta_j(t)$ are negative. The proof is similar to that for Case 2.

Hence, we have proved $L(t+1) \leq a_{\max}L(t) + \delta_3(1-\gamma)^t$ for all $t \geq K_2$.

Denote $b := 1 - \gamma$, then $1 > a_{\max} > b \geq 0$. For any $t \geq K_2$, an upper bound of $L(t)$ is

$$
\begin{aligned}
L(t) &\leq a_{\max}L(t-1) + \delta_3 b^{t-1} \\
&\leq a_{\max}^{t-K_2}L(K_2) \\
&\quad + \delta_3(b^{t-1} + b^{t-2}a_{\max} + \ldots + b^{K_2}a_{\max}^{t-K_2-1}) \\
&= \left(a_{\max}^{-K_2}L(K_2) - \delta_3\frac{b^{K_2}a_{\max}^{-K_2}}{b-a_{\max}}\right)a_{\max}^t + \frac{\delta_3 b^t}{b-a_{\max}}
\end{aligned}
$$

Note that the coefficient of b^t is negative. By choosing δ_4 as the coefficient of $a_{\max}^t$, we get

$$
L(t) \leq \delta_4 a_{\max}^t = \delta_4(1-\gamma+\gamma\mu_{\max})^t
$$

$\square$