Nonlinear Instabilities in TCP-RED

Priya Ranjan, Eyad H. Abed, *Fellow, IEEE*, and Richard J. La

***Abstract*—This work develops a discrete-time dynamical feedback system model for a simplified TCP network with RED control and provides a nonlinear analysis that can help in understanding observed parametric sensitivities. The model describes network dynamics over large parameter variations. The dynamical model is used to analyze the TCP-RED operating point and its stability with respect to various RED controller and system parameters. Bifurcations are shown to occur as system parameters are varied. These bifurcations, which involve the emergence of oscillatory and/or chaotic behavior, shed light on the parametric sensitivity observed in practice. The bifurcations arise due to the presence of a nonlinearity in the TCP throughput characteristic as a function of drop probability at the gateway. Among the bifurcations observed in the system are period doubling and border collision bifurcations. The bifurcations are studied analytically, numerically, and experimentally.**

***Index Terms*—Congestion control, nonlinear instability, RED, TCP.**

I. Introduction

With the growing size and popularity of the Internet, congestion control has emerged as an important problem. Poor management of congestion can render a network partly or fully inaccessible and significantly degrade the performance of networking applications. Researchers have proposed various approaches for addressing this issue. One approach is to keep the network simple and place most of the required intelligence at the end hosts by implementing a more sophisticated end-user rate control allocation [5], [19]. Another approach is to control the congestion level at each router through active queue management (AQM) mechanisms, eg., Random Early Detection (RED) [8], Random Early Marking (REM) [2], Virtual Queue (VQ) [9], and Adaptive Virtual Queue (AVQ) [18]. A common goal of these AQM mechanisms is to detect early signs of congestion and provide feedback to the adaptive sources so that congestion can be avoided without causing a significant degradation in network performance.

The RED mechanism, proposed by Floyd and Jacobson [8], attempts to control the congestion level at a bottleneck by monitoring and updating the average queue size. The basic idea of RED is to sense impending congestion before it happens and provide feedback to the sources by either dropping or marking their packets. The packet marking/dropping probability is the control administered by the RED gateways when they detect queue build-up beyond a certain threshold. Although the RED mechanism is conceptually very simple and easy to understand, its interaction with Transmission Control Protocol (TCP) connections has been found to be rather complex and is not well understood. Most of the rules for setting the parameters of the RED mechanism are based on limited empirical data and come from networking experience. These rules have been evolving, as our understanding of the effects of controller parameters increases. There are reports that discourage wide deployment of RED (e.g., [26]), arguing that there is insufficient consensus on how to select controller parameter values and that RED does not provide a drastic improvement in performance.

Hollot *et al.* have developed a linearized model of a RED gateway with TCP connections to characterize the stability region of the system. Using their single-node analysis, they provided guidelines for selection of RED parameters. Also, based on their observation that exponential averaging introduces an unnecessary feedback delay, they proposed proportional and proportional-integral controllers to improve the responsiveness and stability of the RED mechanism [12], [13].

Low *et al.* have argued that the dynamic behavior of a router queue is governed mostly by the stability of TCP-RED rather than by the detailed dynamics of the Additive-Increase-Multiplicative-Decrease (AIMD) mechanism [23]. This argument forms the basis for multilink multisource utility-based models for studying the interaction of TCP with RED gateways and for a local stability condition in the single-link case. Tinnakornsrisuphap and Makowski used a discrete-time stochastic model to investigate the interaction of a RED gateway with long-lived TCP connections [36]. They have shown that, as the number of TCP connections increases, the queue size per connection converges to a deterministic process, which can be easily computed, and the TCP connections become asymptotically independent. This implies that, with a large number of TCP connections, a RED gateway breaks the synchronization among the connections. Tinnakornsrisuphap, La, and Makowski have extended this model to incorporate the session dynamics, i.e., arrivals and departures of TCP connections, where the workload brought in by a TCP connection is given by a probability law [35]. Similar results have been obtained using differential equations [3].

As noted above, the behavior of a network with TCP at the end nodes and RED at the routers is not well understood and indeed has been found to at times exhibit erratic behavior. To improve the understanding of TCP-RED network dynamics in congested regimes, we follow a *nonlinear* modeling and analysis framework. In congested regimes, when the number of connections is large, the stochastic nature of incoming traffic is of less importance, and the network can be approximated as a deterministic system. This allows us to use nonlinear analysis and

Manuscript received September 29, 2002; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Low. This work was supported in part by the National Science Foundation under Grant ECS 01-5-23927 and Grant ANI 02-37997 and by the Institute for Systems Research at the University of Maryland, College Park.

P. Ranjan is with the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA. He is also with Intelligent Automation, Inc., Rockville, MD 20855 USA.

E. H. Abed is with the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.

R. J. La is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA.

Digital Object Identifier 10.1109/TNET.2004.838600

detailed simulations to explore network behavior over a wide range of parameter values. The major difference between the present work and earlier deterministic studies of TCP-RED is that here we are able to model the generic nonlinear effects of TCP beyond linear operating regimes.

We use a deterministic nonlinear dynamical model of a simple network with TCP connections and a RED gateway. The basic model that we consider was originally proposed by Firoiu and Borden [6]. We modify their model with a simpler TCP throughput function [11], [25] to facilitate the analysis while keeping the dominant nonlinearity. Our work goes beyond a simple linear stability analysis and studies regions where nonlinear instabilities occur due to nonlinearity of the throughput function combined with buffer space limits becoming active. The effect of these nonlinearities and the global dynamics of network protocols have not been explored thoroughly. We show that the model exhibits a rich variety of bifurcation behavior, leading to irregular network operation. As parameters are varied, the system dynamics are shown to transition between a stable fixed point and oscillatory or chaotic behavior. The paper will also describe the detrimental consequences of oscillatory or chaotic behavior on the performance of adaptive sources, such as TCP sources.

We show that the model proposed in [6] can be rewritten as a first-order discrete-time nonlinear dynamical model. This modeling framework is very much in the spirit of self-clocked models proposed by Jacobson [15]. Replacing the TCP throughput function in [6] with a simpler function enables an analytical study of the system and provides us with insight as to how one may be able to improve system behavior. However, our qualitative results do not rely on the particular form of throughput function used.

Finally, this work is directly related to the Network Weather Service [37] and to other similar network performance-forecasting systems generally used in grid computing. The Network Weather Service is a distributed system that periodically monitors and dynamically forecasts the performance that various network and computational resources can deliver over a given time interval. The service operates a distributed set of performance sensors (e.g., network monitors and CPU monitors) from which it gathers readings of instantaneous conditions. It then uses numerical time-series models to generate forecasts of conditions over a given time horizon. The models proposed here and their analysis illustrate how a simple two-node network (client and server) can transition into a regime of unpredictable behavior. The understanding of the underlying dynamics can significantly enhance the prediction capabilities, leading to better QoS and better utilization of networking/computing resources.

The rest of the paper is organized as follows. Section III presents the nonlinear first-order discrete-time model that is used in the analysis. In Section IV, the fixed point of the model is determined and an associated period doubling bifurcation (PDB) is analyzed. The border collision bifurcation from the period-doubled orbit is studied in Section V. Section VI contains numerical examples illustrating bifurcations and nonlinear instabilities in the model. An analytical study of sufficient conditions for chaotic behavior is given in Section VII. In Section VIII, ns-2 simulation results are given. Concluding remarks are collected in Section IX.

II. Background on TCP and RED

In this section, we first briefly explain the main protocols, the interaction of which is the main subject of this paper.

A. Transmission Control Protocol

TCP is the most popular form of congestion control protocol adopted by responsive end-user applications. The transmission rate of a TCP connection is controlled by its congestion window size that determines the maximum number of outstanding packets that have not been acknowledged. TCP operates in two different modes. When a TCP connection is first initiated, it starts in slow start (SS) mode. In SS mode, the connection increases its congestion window size by one for each acknowledged packet until it receives the first congestion notification, e.g., packet drop or marking, at which time it switches to congestion avoidance (CA) mode. In CA mode, the congestion window size is increased by one during the course of a round-trip time (RTT) if the connection does not receive any congestion notification. When a connection receives a congestion notification, the congestion window size is reduced to half of the current value. This is often referred to as the AIMD mechanism.

Due to its popularity and important role in proper management of congestion inside the network, the behavior of TCP has been much studied, and it is well understood in the context of a single flow. It has been shown that, given the RTT R and packet loss probability p, the stationary throughput of a TCP Reno connection can be approximated by

$$T(p,R)=\frac{MK}{\sqrt{p}R}+o\left(\frac{1}{\sqrt{p}}\right) \tag{1}$$

where K is some constant in $[1,\sqrt{8/3}]$ (see [11], [25], [28], and [30]), when packet losses are detected through triple-duplicate acknowledgment. We will use this formula for TCP throughput in our analysis, assuming TCP flows are in CA mode unless mentioned otherwise.

B. Random Early Detection

RED is one of the first AQM mechanisms proposed by Floyd and Jacobson [18]. A RED gateway estimates the congestion level by monitoring and updating its average queue size. In order to maintain a relatively small (average) queue size, rather than waiting until the buffer overflows, it drops a packet with a certain probability to provide an early signal of impending congestion when the average queue size exceeds a threshold. This packet dropping probability p is a function of the average queue size q^{ave} of the following form [18]:[1]

$$p(q^{\text{ave}})=\begin{cases}0, & \text{if } q^{\text{ave}}<q_{\min}\\ 1, & \text{if } q^{\text{ave}}>q_{\max}\\ \dfrac{q^{\text{ave}}-q_{\min}}{q_{\max}-q_{\min}}p_{\max}, & \text{otherwise}\end{cases} \tag{2}$$

[1]In practice, a RED gateway drops a packet with a modified probability in order to lead to a more uniform drop pattern [8].

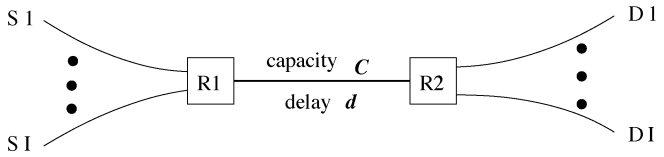

Fig. 1. Topology of the network.

where $q_{\min}$ and $q_{\max}$ are the lower and upper threshold values, and $p_{\max}$ is the selected drop probability when $q^{\text{ave}} = q_{\max}$. The average queue size is updated at the time of packet arrival according to the exponential averaging

$$q_{\text{new}}^{\text{ave}} = (1-w)q_{\text{old}}^{\text{ave}} + w \cdot q_{\text{curr}} \tag{3}$$

where q_{curr} is the current queue size, i.e., the queue size at the time of arrival, and w is the exponential averaging weight, which determines the time constant of the averaging mechanism and how fast the RED mechanism can react to a time-varying load. On the one hand, the averaging weight w should be selected sufficiently small so that transient temporary congestion does not result in an oscillation of the packet drop probability. On the other hand, the averaging weight should be set large enough so that the RED mechanism can react to changes in load in a timely manner. These are two conflicting goals, and the selection of the parameters will affect the interaction of the RED mechanism with adaptive sources, such as TCP. In this paper, however, we show that the averaging weight cannot be set arbitrarily large without causing an oscillatory behavior at the bottlenecks, which affects TCP performance.

III. Discrete-Time Feedback Model for TCP-RED

We consider a simple network where a single bottleneck link is shared by many connections. This is shown in Fig. 1. This can be viewed as a network where there exists a dominant bottleneck shared by connections, e.g., a bottleneck intercontinental Internet link. Let $\mathcal{I}$, $\mathcal{I} = \{1, \cdots, N\}$, denote the set of connections. The set of connections is assumed to remain fixed for the time period of interest. The capacity of the shared link is denoted by C. We assume that the RED queue management mechanism is implemented at each node in order to control the average queue size at the router. If an Explicit Congestion Notification (ECN) mechanism is implemented, the RED gateway marks the packet by setting the ECN bit in the IP header of the packet if the transport layer is ECN capable. This is indicated in the packet through an ECN Capable Transport (ECT) bit in the IP header. If the source is not ECN-capable, the RED gateway drops the packet [7]. The goal of the controller is to keep the aggregate transmission rate of the connections close to the link capacity, while maintaining a reasonably small average queue size between $q_{\min}$ and $q_{\max}$.

All connections are assumed to be long-lived TCP Reno connections. We assume that the connections are uniform and have the same round-trip propagation delay (without any queueing delay), which is denoted by d. However, rather than interpreting this assumption as a requirement that the connections must have the same propagation delay, one should consider the delay d as the effective delay that represents the overall propagation delay of the connections, or this could describe a case where the bottleneck link has a large propagation delay that dominates the round-trip delays of the connections, e.g., an intercontinental Internet link. We denote the rate or throughput of a connection by x and the packet size by M.[2] We also consider a mixture of long-lived and short-lived connections in Section VIII-C.

A network with an AQM mechanism can be modeled as a feedback system, where sources adjust their transmission rates based on feedback from the AQM mechanism in the form of marked or dropped packets [7], [8]. We use a dynamic discrete-time feedback system model, first introduced by Firoiu and Borden [6], to analyze the interaction of a RED gateway with TCP connections.

The control system is defined as follows. At period k, $k = 1, 2, \ldots$, the RED controller at the router provides the feedback signal p_k in the form of a packet drop probability. This feedback signal is a function of the average queue size q_k^{ave} evaluated at period k. Due to a feedback delay introduced by the RTT, the packet drop probability p_k at period k, $k \geq 1$, determines the throughput of the connections and the queue size q_{k+1} at period $k+1$, based on system constraints (such as capacity constraints). The queue size q_{k+1} at period $k+1$ is used to compute the average queue size q_{k+1}^{ave} at period $k+1$ according to the exponential averaging rule in (3). Then, the average queue size q_{k+1}^{ave} is used to calculate the packet drop probability p_{k+1} at period $k+1$, which is the control variable of the AQM mechanism. This can be expressed mathematically as

$$\text{plant function}: \; q_{k+1} = G(p_k) \tag{4}$$

$$\text{averaging function}: \; q_{k+1}^{\text{ave}} = A\left(q_k^{\text{ave}}, q_{k+1}\right) \tag{5}$$

$$\text{control function}: \; p_{k+1} = H\left(q_{k+1}^{\text{ave}}\right) \tag{6}$$

where $A(q_k^{\text{ave}}, q_{k+1})$ is the averaging function

$$A\left(q_k^{\text{ave}}, q_{k+1}\right) = (1-w)q_k^{\text{ave}} + w \cdot q_{k+1} \tag{7}$$

as given in (3), and the RED control function $H(q_{k+1}^{\text{ave}}) = p(q_{k+1}^{\text{ave}})$ as given in (2). The exact form of the plant function $G(\cdot)$ depends on system parameters such as the number N of connections, the nature of the connections, and round-trip delays d. We describe the plant function Section III-A.

Let us first motivate our discrete-time model and explain the relationship with some of the previously proposed models. Since the queue size and average queue size are updated upon packet arrivals, the queue dynamics at a RED gateway evolve at a faster time scale than the RTT of connections. However, because the reaction times of TCP connections are fundamentally limited by their RTTs, the average queue size should not change much over the course of one RTT in order to allow the connections enough time to react to the current level of feedback signal and filter out oscillations due to transient congestion in order for the RED mechanism to work properly, as mentioned earlier. Therefore, the detailed dynamics of the interaction over one RTT will be averaged out by the RED averaging mechanism and will not play a significant role.

[2]For the simplicity of analysis, we assume that all connections use the same packet size. Again, this should be interpreted as the average packet size of the connections rather than a strict requirement.

This observation has been verified in [35] and [36] using a discrete-time stochastic model, where a period is assumed to be an RTT of connections. They show that, as the number of flows becomes large, both queue and average queue sizes converge to deterministic processes (i.e., macroscale model) with details of TCP dynamics filtered out. These results suggest that when modeling a large number of TCP connections the detailed dynamics of interaction with the RED can be simplified using a macroscale model that captures the larger time-scale dynamics roughly at the time scale of RTTs of the connections. Similar results have also been obtained using stochastic differential equations [3]. In addition, our results in [32] illustrate that there exists a natural discrete-time model corresponding to delay-differential equations which arises in studying system stability of a rate-control mechanism with a communication delay, and the stability of one can be inferred from that of the other. In other words, the stability of the system given by delay-differential equations can be determined by studying the stability of the underlying discrete-time system. Therefore, our model here can be interpreted as the underlying discrete-time model corresponding to the system given by delay-differential equations in [3], [12], and [23], which attempt to approximate packet-level dynamics using differential equations.

We have shown [32] that, in case of loss of stability, Slowly Oscillating Periodic (SOP) orbits appear as a result of PDB. These SOP orbits have periods that are somewhat larger than twice the RTT of connections, and the ratio of their periods to the RTT approaches two as the RTT increases.[3] These SOP orbits manifest themselves as period-two orbits in the underlying discrete-time map. This suggests that a natural time scale for studying system stability with a *large* delay is approximately the RTT of connections.

Let us denote the duration of a period in our discrete-time model by T_{period} and the number of packets the link can transmit in a period by n_{period}. Since a period in our model is much larger than typical interarrival times of packets as explained above, the exponential averaging weight in (7) is approximately $w \approx 1-(1-w_{\text{red}})^{n_{\text{period}}} \approx n_{\text{period}} \cdot w_{\text{red}}$ if $w_{\text{red}} \ll 1$, where w_{red} is the exponential averaging weight at a RED gateway.

A. Plant Function

In this subsection, we describe the plant function (4) that will be used for our analysis. In order to compute the plant function, we assume the following. Given the packet drop probability at period k, the aggregate throughput of the connections is given by the stationary throughput formula in (1). In this paper, we follow [28] in taking $K = \sqrt{3/2}$. The exact value of K is not crucial to our analysis. It provides a good approximation for our qualitative results when the RED gateway breaks the synchronization among the connections as demonstrated in [36] and results in relatively uniform packet drops. Moreover, we are mainly interested in characterizing the stability conditions and the parametric sensitivity of system behavior close to the stability region. It has been observed that a system tends to be less stable when the load is light [12], [23], [30], in which case the packet drop rate will be low and the $o(1/\sqrt{p})$ term will be relatively small compared to the first term [28]. Therefore, the first term in (1) will give us an accurate approximation in the region of our interest. This will be validated in Section VIII through *ns*-2 simulation.

We use this simple approximation for TCP throughput to facilitate our analysis. The use of a stationary throughput formula for computing the plant function can be justified as follows. Since the queue size changes continually between two consecutive periods, i.e., over the duration of a period, and the feedback signal also varies continually, the connections see the time-varying feedback signal from the gateway and can react to the change in a continuous manner rather than having to wait till the end of next RTT. When the system is stable, the average queue size will remain close to the equilibrium queue size when the number of connections is large, as proved in [3] and [36]. Thus, the *aggregate* throughput of the connections will see the stationary throughput. Therefore, this assumption is clearly valid while the system remains stable. When the system becomes unstable, it has been observed in [16] and [32] that the connections tend to become synchronized. Therefore, if the exponential averaging weight is reasonably small, as it should be, and the number of connections is large with nonnegligible RTT, the *aggregate* throughput of connections will be able to adapt to the *slow* time-varying feedback signal in a timely manner, and it will take only a few RTTs for the set of connections to see the effects introduced by a small change in drop probability. We will illustrate this using a numerical example in Section VI, where a period in the discrete-time model is shown to be roughly three times the mean RTT of connections. Similar behavior has also been observed in [32]. The fact that connections continuously react to time-varying feedback signal in practice suggests that the dynamics of connections are likely to be slightly smoother in practice than our model would predict.

In addition, our qualitative results do not depend on this particular form of TCP throughput approximation and are consequences of the nonlinear dependence of TCP throughput on drop probability p, of which (1) is one instance. Similar results to those obtained here hold for more detailed TCP throughput function models.

Since the aggregate throughput of connections cannot be larger than the link capacity, this determines the queue size at the next period $k+1$ as follows. First, we can compute the steady-state packet drop probability p_u such that the bandwidth capacity constraint is satisfied as follows:

$$\sum_{i \in \mathcal{I}} T(p_u, d) = N \cdot \frac{MK}{\sqrt{p_u} d} = C. \tag{8}$$

This is the smallest probability that results in a queue size of zero at the next period, and, for all $p_k > p_u$, the queue size is zero at the next period. Hence, if $p_k \geq p_u$, we know that the throughput of the TCP connections is given by $(MK/(\sqrt{p_k} d))$ and the queue size at period $k+1$ is zero, i.e., $q_{k+1} = 0$. From (8), we can derive that

$$p_u = \left(\frac{NMK}{dC} \right)^2 \tag{9}$$

[3]The exact ratio of the period of the SOP orbits to the RTT depends on the system parameters and could be considerably larger than two if the RTT is small.

and the corresponding average queue size q_u^{ave} such that for any $q_k^{\text{ave}} \geq q_u^{\text{ave}}$, q_{k+1} is identically zero is given by

$$q_u^{\text{ave}} = \begin{cases} \frac{p_u(q_{\max}-q_{\min})}{p_{\max}} + q_{\min}, & \text{if } p_{\max} \geq p_u \\ q_{\max}, & \text{otherwise.} \end{cases}$$

Suppose first that the buffer size B is infinite. If $p_k < p_u$, the bottleneck link capacity is fully utilized. Thus, if $p_k < p_u$, one can obtain the queue size q_{k+1} at period $k+1$ as the solution of the following equation:

$$\frac{MK}{\sqrt{p_k}\left(d + \frac{q_{k+1} \cdot M}{C}\right)} = \frac{C}{N}. \tag{10}$$

The interpretation of (10) is as follows. Assuming symmetric TCP connections, the bottleneck link capacity is equally divided among the TCP connections. In this case, the throughput of a TCP connection will be given by $T(p_k, R(q_{k+1})) = (MK/(\sqrt{p_k}(d + (q_{k+1}M/C)))) = C/N$. Hence, the queue occupancy q_{k+1} is given by

$$q_{k+1} = \frac{C}{M}\left(\frac{MKN}{\sqrt{p_k}C} - d\right). \tag{11}$$

Now let the buffer be of finite size B. From (11), we see that q_{k+1} is a strictly decreasing function of p_k, and hence we can compute the largest p_k such that the queue size q_{k+1} equals the buffer size B. This probability, which we denote by p_l, is given by $(NMK/(dC + BM))^2$. The corresponding average queue size q_l^{ave} is

$$q_l^{\text{ave}} = \frac{p_l(q_{\max} - q_{\min})}{p_{\max}} + q_{\min}.$$

It is obvious that, for all $p_k \leq p_l$, i.e., $q_k^{\text{ave}} \leq q_l^{\text{ave}}$, we have $q_{k+1} = B$. From (8) and (11), we have the full definition of the plant function

$$\begin{aligned} G(p_k) &= \begin{cases} 0, & \text{if } p_k \geq p_u \\ B, & \text{if } p_k \leq p_l \\ \frac{NK}{\sqrt{p_k}} - \frac{Cd}{M}, & \text{otherwise} \end{cases} \\ &= q_{k+1}. \end{aligned} \tag{12}$$

This type of plant function has been verified by ns-2 simulation by Firoiu and Borden [6] using a particular TCP throughput function similar to that used here.

From (4)–(6) and (12), we obtain the mapping

$$\begin{aligned} q_{k+1}^{\text{ave}} &= (1-w)q_k^{\text{ave}} + w \cdot A(G(H(q_k^{\text{ave}}))) \\ &= \begin{cases} (1-w)q_k^{\text{ave}}, & \text{if } q_k^{\text{ave}} \geq q_u^{\text{ave}} \\ (1-w)q_k^{\text{ave}} + w \cdot B, & \text{if } q_k^{\text{ave}} \leq q_l^{\text{ave}} \\ (1-w)q_k^{\text{ave}} + w \cdot \left(\frac{NK}{\sqrt{p_k}} - \frac{Cd}{M}\right), & \text{otherwise} \end{cases} \\ &:= g(q_k^{\text{ave}}, \rho) \end{aligned} \tag{13}$$

where ρ summarizes the system parameters, including the exponential averaging weight w, and $p_k = ((q_k^{\text{ave}} - q_{\min})/(q_{\max} - q_{\min}))p_{\max}$ from (2). This mapping gives the dynamical relationship of the average queue size at period $k+1$ to the average queue size at period k, as shown in Fig. 2. There are three segments in this map showing either increasing and decreasing behaviors of average queue size in different regimes. Most of the interesting dynamics occur due to the middle segment of the map. There are two types of forces in this segment, one of which arises from averaging and the other arises from the RED control action. Their relative contributions in the queue occupancy in the next period are determined by the averaging parameter w. This interaction of averaging and RED control law is crucial to the kind of instabilities and instability cascades that occur as a system or RED parameter is slowly varied.

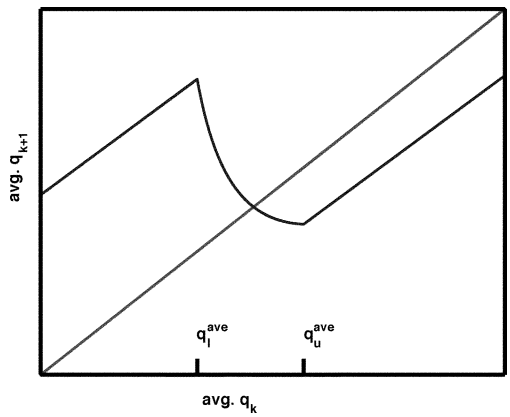

Fig. 2. First return map for TCP-RED.

IV. Fixed Point and Its Bifurcation

A fixed point of the mapping $g(\cdot)$ is an average queue size q^* such that $q^* = g(q^*, \rho)$. If the RED parameters are properly configured, then the average queue size should remain between $q_{\min}$ and $q_{\max}$.

Assumption 1: $p_{\max} > p_u$, where p_u is the largest probability that yields the full utilization, defined in (9). This is natural from a practical point of view since it disallows a disconnected RED law wherein the drop probability jumps from $p_{\max}$ to 1.

Under Assumption 1, solving (13) for a fixed point q^* leads to a third-order polynomial, which does not depend on the exponential averaging weight w because neither the "queue law" nor the "feedback control law" is a function of w. The corresponding probability p^* of the fixed point q^* is given as the square of the positive real solution of the polynomial

$$\frac{CM}{\nu}y^3 + (CMq_{\min} + dC^2)y - NMKC = 0 \tag{14}$$

where $\nu = (p_{\max}/(q_{\max} - q_{\min}))$.

The linear stability of the fixed point q^* can be studied by considering the associated eigenvalue

$$\begin{aligned} \frac{\partial}{\partial q_k^{\text{ave}}} g(q_k^{\text{ave}}, \rho)\Big|_{q^*} &= 1 - w - \frac{wNK}{2\sqrt{\nu}(q^* - q_{\min})^{\frac{3}{2}}} \\ &:= \lambda(q^*(\rho), \rho) \end{aligned} \tag{15}$$

The linear stability condition is $|\lambda(q^*(\rho), \rho)| < 1$ or

$$\left|1 - w - \frac{wNK}{2\sqrt{\nu}(q^* - q_{\min})^{\frac{3}{2}}}\right| < 1. \tag{16}$$

In order to simplify the analysis, we reduce the number of parameters in the model by performing a normalization.

A. Normalization Scheme

Define the parameter γ as

$$\gamma := \frac{q_{\max} - q_{\min}}{p_{\max} B} = \frac{1}{\nu B} > 0.$$

The normalized state variables and RED queue thresholds are defined as

$$\begin{aligned}
q_k^n &:= \frac{q_k^{\text{ave}}}{B}, \quad q_{\min}^n := \frac{q_{\min}}{B} \\
q_l^n &:= \frac{q_l^{\text{ave}}}{B} = \left(\frac{NK}{B + \frac{dC}{M}}\right)^2 \gamma + q_{\min}^n \\
q_u^n &:= \frac{q_u^{\text{ave}}}{B} = p_u \cdot \gamma + q_{\min}^n \\
q_{k+1}^n &= \begin{cases} (1-w)q_k^n, & \text{if } q_k^n > q_u^n \\ (1-w)q_k^n + w, & \text{if } q_k^n < q_l^n \\ (1-w)q_k^n & \\ \quad + w\left(\frac{NK}{B\sqrt{\frac{q_k^n - q_{\min}^n}{\gamma}}} - \frac{dC}{MB}\right), & \text{otherwise} \end{cases} \\
&:= f(q_k^n, \rho). \qquad (17)
\end{aligned}$$

Equation (17) maps the unit interval into itself.

B. Bifurcation Analysis

Local stability of a one-dimensional (1-D) map in the neighborhood of a fixed point is determined by the eigenvalue of the linearized map evaluated there. For the normalized map (17), this eigenvalue is

$$\begin{aligned}
\left.\frac{\partial f(q_k^n, \rho)}{\partial q_k^n}\right|_{q_n^k = q^{n*}} &= 1 - w - \frac{wNK}{2B(q^{n*} - q_{\min}^n)^{\frac{3}{2}}}\sqrt{\gamma} \\
&:= \lambda(\rho) \qquad (18)
\end{aligned}$$

where q^{n^*} is the fixed point of the normalized map. The linear stability criterion ($|\lambda(\rho)| < 1$ [14]) is

$$\left|1 - w - \frac{wNK}{2B(q^{n*} - q_{\min}^n)^{\frac{3}{2}}}\sqrt{\gamma}\right| < 1. \qquad (19)$$

Note that the eigenvalue depends on the fixed point. Of significant interest here are the parameter settings which may lead to loss of stability of the fixed point, giving rise to nonlinear instabilities through a system bifurcation. Numerical simulations of the system show the presence of oscillatory regimes as control and system parameters are varied and indicate that a PDB occurs from the fixed point with the variation of any of the system or control parameters. Thus, we are led to consider cases in which the eigenvalue given by (18) becomes -1, giving a PDB leading to oscillatory behavior in the system. To demonstrate the existence of such bifurcations, it is easiest to focus on the exponential averaging parameter w as the distinguished bifurcation parameter. The critical value of w is one for which the eigenvalue given by (18) is -1. The critical value can be expressed in a closed form as follows:

$$w_{\text{crit}} = \frac{2}{1 + \frac{nK}{2(q_e^* - q_{\min})^{\frac{3}{2}}}\sqrt{\frac{q_{\max} - q_{\min}}{p_{\max}}}} \qquad (20)$$

where q_e^* is a fixed point of the system whose corresponding probability is given as a square of the solution from (14).

PDB can be supercritical or subcritical. In the supercritical case, attracting period-two orbits emerge from the fixed point on the unstable side of the fixed point. In the subcritical case, repelling period-two orbits emerge on the stable side. The ramifications of these two types of PDB for system behavior are very different, with supercritical bifurcation leading to a steady oscillatory behavior near the original fixed point and subcritical bifurcation leading to divergent oscillations. It is possible to determine analytically which of these two cases will arise [14]. To do so, we need to compute the second and third derivatives of the normalized map

$$\left.\frac{\partial^2 f}{\partial q_k^{n2}}\right|_{q_n^k = q^{n*}} = \frac{3wNK}{4B(q^{n*} - q_{\min}^n)^{\frac{5}{2}}}\sqrt{\gamma} \qquad (21)$$

$$\left.\frac{\partial^3 f}{\partial q_k^{n3}}\right|_{q_n^k = q^{n*}} = \frac{-15wNK}{8B(q^{n*} - q_{\min}^n)^{\frac{7}{2}}}\sqrt{\gamma} \qquad (22)$$

to analyze the nature of this bifurcation. The quantity $S = ((1/2)(\partial^2 f/\partial q_k^{n2})^2 + (1/3)(\partial^3 f/\partial q_k^{n3}))$ (evaluated at the fixed point and the selected parameter values) determines the nature of a PDB (see [10, p. 158]. A positive S implies that the bifurcation is supercritical, and a negative S implies a subcritical bifurcation. For the system given by (17), S is

$$S = \frac{wNK\sqrt{\gamma}}{B(q^{n*} - q_{\min}^n)^{\frac{7}{2}}}\left[\frac{9}{32}\frac{wNK\sqrt{\gamma}}{B(q^{n*} - q_{\min}^n)^{\frac{3}{2}}} - \frac{5}{8}\right]. \qquad (23)$$

The expression for S in (23) shows that it may change sign giving rise to a subcritical bifurcation if the parameters are in certain ranges. This should be kept in mind when designing a TCP-RED system to avoid any unexpected oscillations in router queues.

First, suppose that the system and control parameters are fixed, except for the averaging weight w. Then, from (15), we see that the eigenvalue is a linearly decreasing function of w, becoming more negative as w is increased. Now consider the critical averaging weight w^* to be a function of N and denote it as $w^*(N)$. Then

$$w^*(N) = \frac{2}{1 + \frac{NK}{2\sqrt{\nu}(q^* - q_{\min})^{\frac{3}{2}}}}.$$

The next lemma states that the largest value of the averaging weight that can be used without resulting in loss of stability is an increasing function of N.[4]

Lemma 1: The critical parameter value $w^*(N)$ is an increasing function of N.

A proof of this lemma can be found in [21]. This lemma tells us that, when the load is light, the averaging weight must be selected small in order to avoid an oscillatory behavior in the queue size due to a PDB. The importance of the bifurcation point is that the system quickly becomes very unstable in the sense that the queue size oscillates widely, often resulting in an empty queue, reducing the system throughput and increasing the RTT variance of TCP connections. This will be shown in

[4]For the rest of the paper, we limit our interests to the region where $q^* \leq q_{\max}$.

Sections VI and VIII. One can show in a similar manner that the initial PDB point $w^*(\cdot)$ is a decreasing function of the round-trip propagation delay d and $q_{\min}$ and an increasing function of $q_{\max}$ when these parameters are varied in isolation while other parameters are fixed.

Below, some analytical properties of the map (17) are given and proved, in preparation for the study of possible instability routes in the next section.

Assumption 2: Assume that the left derivative of the normalized map in (17) is negative for $q_k^n = q_u^n$.

This assumption is not very restricting. It simply asserts that, as the (normalized) average queue size q_k^n increases from q_l^n to q_u^n, the average queue size in the next period computed according to the map in (17) decreases with q_k^n. This assumption will be true if the negative feedback component of the RED is larger than the contribution retained by the averaging mechanism in the middle segment of the map shown in Fig. 2.

Lemma 2: The map given by (17) is piecewise monotone under Assumptions 1 and 2.

Next, we analyze the parametric dependence of the TCP-RED system and show that it is smooth in parameter with respect to w.

Lemma 3: The map given by (17) depends smoothly on w.

We refer the reader to [29] for a proof of Lemmas 2 and 3. Properties of TCP-RED map outlined by these lemmas will be used in the next section to leverage border collision bifurcation theory to the understand the dynamics in different regions.

V. Border Collision Bifurcation

In this section we use the *border collision bifurcation* (BCB) theory [4], [24] to analyze the bifurcations due to the variation of parameter w. BCBs occur for piecewise smooth maps and involve a nonsmooth bifurcation occurring when a parameter change results in a fixed point (or other operating condition) crossing a border between two regions of smoothly defined dynamics in state space.

If a fixed point collides with the border(s) with a change in the parameters, there is a discontinuous change in the derivative $\partial f/\partial x$ of map $f(x)$, and the resulting phenomenon is called border collision bifurcation. This kind of bifurcation has been reported widely in economics [27], mechanical systems, and power electronic models [4], [24], [27].

Border collision is a local bifurcation, and hence it can be studied by characterizing the local properties of a map in the neighborhood of the colliding border. It is shown in [27] that a normal form which is an affine approximation of f in the border neighborhood is sufficient to quantify the possible BCBs. This normal form is

$$G(x,\mu) = \begin{cases} ax + \psi, & \text{if } x \le 0 \\ bx + \psi, & \text{if } x \ge 0 \end{cases} \tag{24}$$

where

$$a = \lim_{x \to x_b^-} \frac{\partial}{\partial x} f(x, \psi^*), \quad b = \lim_{x \to x_b^+} \frac{\partial}{\partial x} f(x, \psi^*) \tag{25}$$

and ψ^* is the parameter for which border collision happens. It can be assumed to be 0 without any loss of generality.

There are various types of bifurcation scenarios possible depending on the values of coefficients a and b in the normal form given in (25). For the sake of simplicity, we will discuss only the case relevant to the observed phenomena in our model and provide a numerical proof by computing the one-sided coefficients (eigenvalues) for the same.

The following lemma from [4] shows that the border has a crucial role if a certain bifurcation sequence occurs.

Lemma 4: If a fixed point of the map given by (17) undergoes a smooth $(eigenvalue = -1)$ PDB at w_1 and the resulting period-two orbit also goes through a smooth period doubling for $w_2 > w_1$, then, under the piecewise monotonicity condition, the periodic orbit must collide with the border for some $w \in [w_1\ w_2]$.

We will see this kind of smooth and nonsmooth bifurcation in the next section when we present numerical examples.

For our model, the case of interest in border collision theory is when

$$0 < a < 1 \text{ and } b < -1. \tag{26}$$

This is mentioned in [27, case 8]. It is shown that, in this case, a fixed point attractor can bifurcate into a periodic attractor or a chaotic attractor as ψ is varied from negative to positive. This is the exact phenomenon we observe for our model when the bifurcation parameter w is varied and a stable period-two orbit transitions to chaos. Essentially, if we take the second iterate of our map, it exhibits a fixed point bifurcating into a chaotic orbit. The existence of chaos can be confirmed by computing the Lyapunov exponents [30]. A numerical example that provides evidence for our claim is given in the next section.

VI. Numerical Examples

The behavior of the map in (13) can be explored numerically in parameter space to look for interesting dynamical phenomena. When the eigenvalue exits the unit circle, the fixed point becomes unstable. Depending on the nature of the ensuing bifurcation, there can be new fixed points, higher period orbits, or chaos. There is also a possibility of an orbit (original fixed point or a bifurcated orbit) colliding with either border q_u^{ave} or q_l^{ave}, leading to a rich set of possible bifurcations.

In this section, we numerically validate our analysis using bifurcation diagrams. A bifurcation diagram shows the qualitative changes in the nature and the number of fixed points of a dynamical system as parameters are quasi-statically varied. The horizontal axis is the parameter that is being varied, and the vertical axis represents a measure of the steady states (fixed points or higher period orbits). For generating the bifurcation diagrams, in each run we randomly select four initial average queue sizes, $q_1^{\text{ave}}(0)$, $q_2^{\text{ave}}(0)$, $q_3^{\text{ave}}(0)$, and $q_4^{\text{ave}}(0)$, and these average queue sizes evolve according to the map $g(\cdot)$ in (13), i.e.,

$$\begin{gathered} q_i^{\text{ave}}(k) = g\left(q_i^{\text{ave}}(k-1), \rho\right), \\ \text{for } k = 1, \cdots, 1{,}000 \\ \text{and } i = 1,\ 2,\ 3,\ \text{and } 4. \end{gathered}$$

We plot $q_i^{\text{ave}}(k)$, $k = 991, \cdots, 1{,}000$, and $i = 1, 2, 3$, and 4. Hence, if there is a single stable fixed point or attractor q^* of the system at some value of the parameter, all $q_i^{\text{ave}}(k)$ will converge to q^* and there will be only one point along the vertical line at the value of the parameter. However, if there are two stable fixed

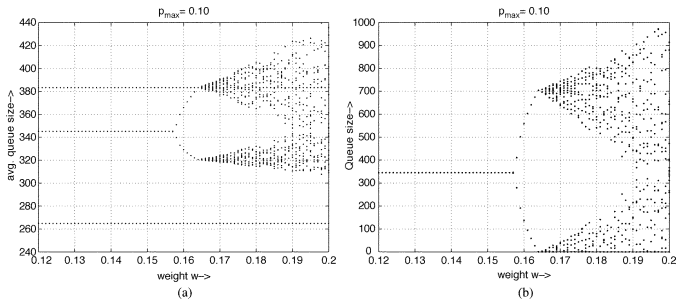

Fig. 3. Bifurcation diagram of average and actual queue length with respect to the averaging weight w ($p_{\max} = 0.1$).

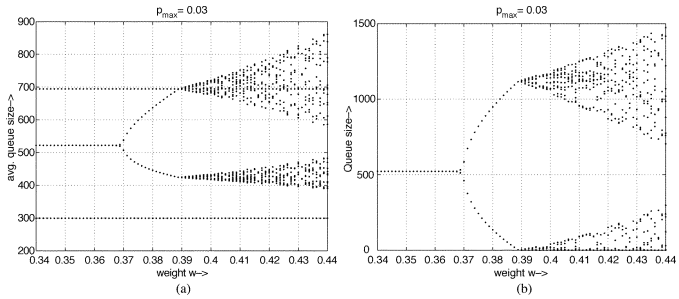

Fig. 4. Bifurcation diagram of average and actual queue length with respect to the averaging weight w ($p_{\max} = 0.03$).

points, $\tilde{q}_1^{\text{ave}}$ and $\tilde{q}_2^{\text{ave}}$, with a period of two, i.e., $g(\tilde{q}_i^{\text{ave}}, \rho) \neq \tilde{q}_i^{\text{ave}}$ and $g(g(\tilde{q}_i^{\text{ave}}, \rho)) = \tilde{q}_i^{\text{ave}}$, $i = 1, 2$, then there will be two points along the vertical lines and the average queue size will alternate between $\tilde{q}_1^{\text{ave}}$ and $\tilde{q}_2^{\text{ave}}$.

In this section, we study the effects of various system and control parameters on average and queue behavior as each of these parameters is varied while the others are fixed. More specifically, we study how the averaging weight w, lower threshold $q_{\min}$, the number of connections N, and the round-trip propagation delay d affect system stability, queue behavior, and their sensitivity to these parameters.

A. Effect of Exponential Averaging Weight

We use the following parameters for the numerical examples presented in this subsection:

$$q_{\max} = 750,\ q_{\min} = 250,\ C = 75 \text{ Mb/s},$$
$$K = \sqrt{\frac{3}{2}},\ B = 3,750 \text{ packets},\ M = 4,000 \text{ b},$$
$$N = 250,\ d = 0.1 \text{ s},\ w = \text{bifurcation parameter}.$$

The bifurcation plots in Figs. 3 and 4 show the effect of varying the averaging weight w for different values of $p_{\max}$, namely $p_{\max} = 0.1$ and $p_{\max} = 0.03$. Figs. 3(a) and 4(a) show the exponentially averaged queue sizes, and Figs. 3(b) and 4(b) plot the actual queue sizes. For small w, these plots have a fixed point, which shows up as a straight line until some critical value of w is reached, at which point the straight line splits into two. The emergence of two stable fixed points of period two is a consequence of a PDB. This is the first indication of oscillatory behavior appearing in the system due to the inherent nonlinearity of the interaction between RED mechanism and TCP, as opposed to a discontinuity in "queue or control law" which has been suggested in the past. This period-two oscillation starts batching load at the router, as shown in the plots.

TABLE I
EIGENVALUES COMPUTED FOR DIFFERENT VALUES OF PARAMETER w TO ILLUSTRATE PDB, $q_u^{\text{ave}} = 0.102222$

w	q_k^n	$\lambda(q_k^n, w)$	Legend
0.1561	0.092028	-0.978111	Close to PDB
0.1572	0.092028	-0.992051	Closer to PDB
0.1583	0.092028	-1.005990	After PDB
0.1594	0.092028	-1.019929	After PDB

TABLE II
EIGENVALUES COMPUTED FOR DIFFERENT VALUES OF PARAMETER w TO ILLUSTRATE BCB, $q_u^n = 0.102222$

w	q_k^n	q_{k-1}^n	q_{k-2}^n	q_{k-3}^n
0.1620	0.100108	0.086412	0.100108	0.086412
0.1631	0.085846	0.101330	0.085846	0.101330
0.1642	0.102193	0.085488	*0.102283*	0.085447

w	$\lambda^2(k, k-1)$	$\lambda^2(k-2, k-3)$	Legend
0.1620	0.786415	0.786415	Before BCB
0.1631	0.729421	0.729421	Before BCB
0.1642	0.692157	-1.815238	After BCB

Increasing w further results in one of the period-two fixed points colliding with the upper border of the map, giving a chaos-type phenomenon. This is basically a bifurcation sequence expressed briefly as $1 \rightarrow 2 \rightarrow chaos$. This is a case of BCB, as shown in the analysis earlier. It can be seen that, when the bifurcation diagram for q_k^{ave} collides with the border q_u^{ave}, the queue empties, underutilizing the bottleneck link capacity. The implication of a relatively small oscillation in the average queue length is rather serious for the queue length since the buffer starts getting empty and overly filled in every alternate cycle. This dynamical phenomenon is common to both plots in Figs. 3 and 4. We note that the distance between the initial PDB point and the BCB point is short in both cases. This suggests that an effective way of controlling the instability may be to control the first PDB point, as demonstrated in [21] and [33].

To illustrate the PDB in the system, we compute the eigenvalue for the fixed point as w is varied. It can be seen that this eigenvalue leaves unit circle along a negative real line, indicating a PDB. We also track the unstable fixed point and compute the corresponding eigenvalue to show that it indeed crosses the unit circle as shown in Table I. We also notice that both the stable and unstable fixed point ($q^{n*} = 0.092028$) is smaller than $q_u^{\text{ave}} = 0.102222$ for the normalized model. Hence, it lies on the same side of the border even after smooth PDB.

To provide evidence for our claim for a BCB, we further compute the eigenvalue of a period-two orbit of the map numerically and show that, indeed, one-sided eigenvalues obey the condition given in (26). This computation is done for the set of parameters corresponding to Fig. 3. We define $\lambda^2(i, j) = \lambda(q_i^n) * \lambda(q_j^n)$.

In Table II, the first and second rows show the four consecutive states (the exponentially averaged queue size at the router) corresponding to the parameter w just before the BCB but after PDB. We note that all the states stay on the same side of the border with eigenvalue corresponding to a period-two orbit being less than unity. This implies the existence of stable period-two orbit.

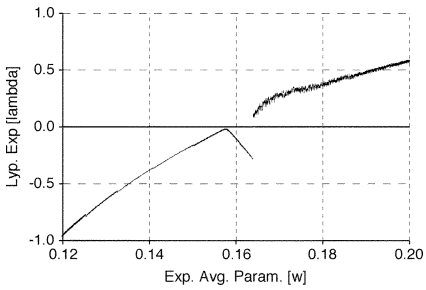

Fig. 5. Lyapunov exponent computed for average queue length with respect to the averaging weight w ($p_{\max} = 0.1$).

The third row depicts the same data just after a BCB from a fixed point to chaos for the second iterate. Comparing the states with the border (q_u^n) reveals that q_{k-2}^n and q_{k-3}^n lie on different sides of the border. The eigenvalues corresponding to these two points, i.e., $\lambda^2(k-2, k-3)$, is negative. This eigenvalue $\lambda^2(k-2, k-3)$ can be used to approximate b in (25) in this case. Similarly, the eigenvalue corresponding to q_{k-1}^n and q_k^n, i.e., $\lambda^2(k, k-1)$, can be used to approximate a in (25). Since a lies between 0 and 1, as shown in Table II, and b is smaller than -1, these values satisfy the condition given by (26). Note that the eigenvalues change discontinuously as w is varied. This supports our contention that there is a *border collision* bifurcation in the system through which the system may become chaotic. It also stresses the role played by a border. We also note that there is a possibility of other rich nonlinear instabilities with different periodicity based on different parameter settings.

We also plot the Lyapunov exponents for the bifurcation scenario in Fig. 3 where $p_{\max} = 0.1$. This is useful since a positive Lyapunov exponent indicates the presence of chaotic behavior [1, p. 110]. Fig. 5 shows that for small w the exponent is negative, which corresponds to the single stable fixed point. It slowly increases to zero near the PDB and then becomes negative again due to a stable period-two orbit. Finally, it jumps to a positive value when one of the period-two fixed points collides with one of the borders.

B. Effect of Lower Threshold Value

In this subsection, we explore how the lower threshold value $q_{\min}$ affects system stability and behavior. The set of parameters used for the numerical example in this section is given as follows:

$$q_{\max} = 750,\ \ w = 0.15,\ \ C = 75 \text{ Mb/s},\ \ K = \sqrt{\frac{3}{2}},$$
$$B = 3,750 \text{ packets},\ \ d = 0.1 \text{ sec},\ \ M = 4,000 \text{ b},$$
$$N = 250,\ \ p_{\max} = 0.1,\ \ q_{\min} = \text{bifurcation parameter}.$$

Fig. 6(a) shows the fixed points of the system with varying $q_{\min}$. As is clear from the figure, similar nonlinear phenomena are exhibited in this case as well. The system becomes less stable as $q_{\min}$ is increased, while other parameters are held fixed. Furthermore, this plot also exhibits similar sensitivity of the system behavior to $q_{\min}$, as in the previous subsection.

We plot the Lyapunov exponent corresponding to the bifurcation scenario in Fig. 6(a). The Lyapunov exponent shown in

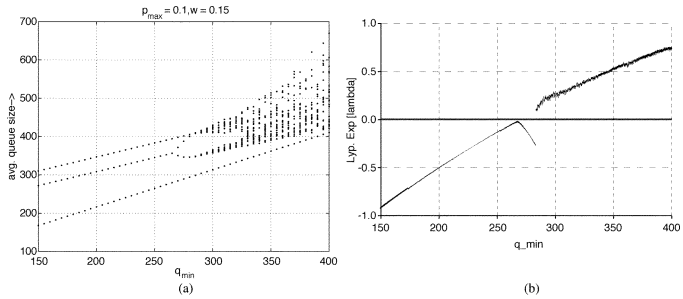

Fig. 6. (a) Bifurcation diagram of average and actual queue length with respect to $q_{\min}$ ($p_{\max} = 0.1, w = 0.15$). (b) Lyapunov exponent computed for average queue length with respect to $q_{\min}$.

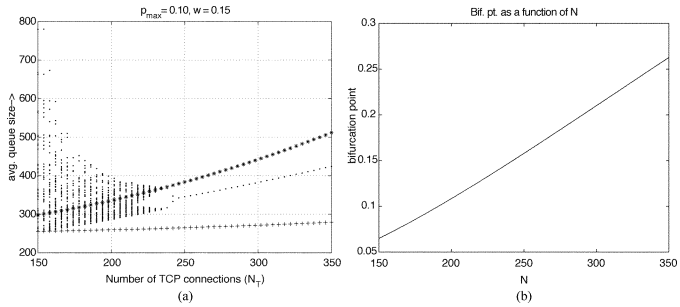

Fig. 7. (a) Bifurcation diagram of average queue length with respect to the number of connections N ($p_{\max} = 0.1, w = 0.15$). (b) Initial PDB point as a function of a number of active TCP-sessions.

Fig. 6(b) also stays negative for small $q_{\min}$ as in Fig. 5. In a similar fashion, it increases to zero when the system goes through a PDB and again decreases when the system has a stable period-two trajectory. Finally, it jumps to a positive value after a BCB and stays positive.

C. Effect of System Parameters on Stability

A network manager may have control over the selection of control parameters such as w, $q_{\min}$, $q_{\max}$, and $p_{\max}$. However, other system parameters such as the number of connections N and the round-trip propagation delay d are out of the network manager's control. Hence, it is important to understand how these system parameters affect the system stability and queue behavior in relation to the control parameters in order to understand how these control parameters should be set in practice.

We first study the impact of the number of connections N on the system behavior. The system parameters used for our example are as follows:

$$q_{\max} = 750,\quad q_{\min} = 250,\quad C = 75 \text{ Mb/s},\quad K = \sqrt{\frac{3}{2}},$$
$$B = 3,750 \text{ packets},\quad d = 0.1 \text{ s},\quad M = 4,000 \text{ b},$$
$$w = 0.15,\quad p_{\max} = 0.1,\quad N = \text{bifurcation parameter}.$$

The bifurcation diagram in Fig. 7(a) shows that the system stabilizes as the number of connections N increases. In general, there is an agreement that a larger number of users tends to stabilize the system [12], [23].

Another way of verifying that the system stability improves with N is to compute the initial PDB point, i.e., the averaging

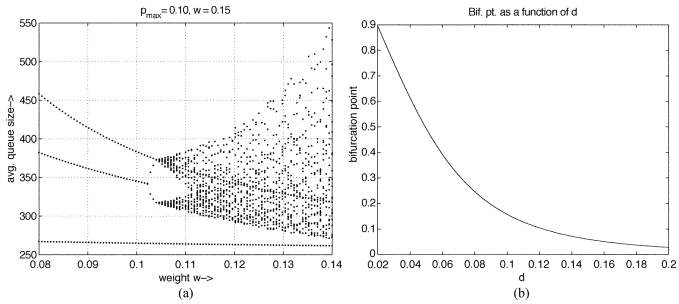

Fig. 8. (a) Bifurcation diagram of average and actual queue length with respect to round-trip propagation delay (d). (b) Initial PDB point as a function of round-trip propagation delay.

weight w^* at which the PDB occurs, as a function of N. This is shown in Fig. 7(b). The plot shows that the critical bifurcation parameter value $w^*(N)$ increases with the number of active TCP connections. Its implication for stability is that increasing the number of active TCP connections renders the queue length stable, though at the expense of increased delay, since a larger value of w is needed to induce the first PDB and consequently BCB. This is in agreement with results in [12] and [23], where under certain conditions a larger number of active TCP sessions stabilizes the system.

Similarly, we also plot a bifurcation diagram with respect to round-trip propagation delay d. The plot in Fig. 8(a) is in general agreement with the result in [12] and [23] that larger delays cause instability. Parameters for this bifurcation diagram are given as follows:

$$q_{\min} = 250, \quad q_{\max} = 750, \quad C = 75 \text{ Mb/s}, \quad K = \sqrt{\frac{3}{2}},$$
$$B = 3,750 \text{ packets}, \quad N = 250, \quad M = 4,000 \text{ b},$$
$$w = 0.15, \quad p_{\max} = 0.1, \quad d = \text{bifurcation parameter}.$$

The variation of the critical value of the bifurcation parameter $w^*(d)$ as a function of round-trip propagation delay d is plotted in Fig. 8(b). It shows that the system is more stable for smaller values of round-trip propagation delay, as a larger averaging weight w is needed to make it oscillate. This result again is in agreement with the general result of [12] and [23] that smaller delays tend to keep the system stable. One should note that the system stability is rather sensitive to the variation in d.

VII. Chaotic Behavior

The purpose of this section is to give an analytical proof of the presence of chaos in the TCP-RED dynamic model. The tool we use is a well-known theorem of Sharkovsky [34], which was also proved by Li and Yorke [22] and goes by the name "period three implies chaos." It applies to continuous 1-D maps and thus can be applied to piecewise-smooth but continuous systems such as the system studied here.

The main result of Li and Yorke [22] is as follows.

Theorem 1: Let J be an interval and let $F : J \to J$ be continuous. Assume that there is a point $a' \in J$ for which the points $b' = F(a')$, $c' = F^2(a')$, and $d' = F^3(a')$ satisfy

$$d' \le a' < b' < c' \text{ or } d' \ge a' > b' > c'.$$

Then

1) for every $k = 1, 2, \ldots$ there is a periodic point in J having period k;
2) there is an uncountable set $S \subset J$ (containing no periodic points), which satisfies the following conditions.

a) For every $p, q \in S$ with $p \ne q$,

$$\limsup_{n\to\infty} |F^n(p) - F^n(q)| > 0$$

$$\liminf_{n\to\infty} |F^n(p) - F^n(q)| = 0.$$

b) For every $p \in S$ and periodic point $q \in J$,

$$\limsup_{n\to\infty} |F^n(p) - F^n(q)| > 0.$$

In our case $J = [0\ B]$, and F is given by the function $g(\cdot,\cdot)$ which defines the TCP-RED map in (13). It is also clear that the TCP-RED map is continuous by construction as long as Assumption 1 is in force. Also, note that the existence of a period-three orbit, i.e., $d' = a' > b' > c'$ or $d' = a' < b' < c'$, is a special case of the hypotheses of the theorem and proves the existence of chaos.

We have proved earlier [29] that the map (13) is strictly increasing for $0 \le q^{\text{ave}} \le q_l^{\text{ave}}$ and for $q_u^{\text{ave}} \le q^{\text{ave}} \le B$, but it can be strictly decreasing in the segment where $q_l^{\text{ave}} \le q^{\text{ave}} \le q_u^{\text{ave}}$ under certain conditions.

To apply the theorem, we need to choose a starting point a' and iterate on it using the map g three times and then apply the conditions stated in the theorem. We select $a' = (q_u^{\text{ave}}/(1-w))$. This choice is made based on earlier numerical studies (Matlab) which showed a strong tendency toward bifurcation and chaos when the system state q_k^{ave} nears q_u^{ave}. With this choice for a', we find that $b' = g(a') = (1-w)a' = q_u^{\text{ave}}$ and $c' = g^2(a') = g(q_u^{\text{ave}}) = (1-w)q_u^{\text{ave}}$. Looking at Fig. 2, it is clear that there are two possible cases for the location of c': either $q_l^{\text{ave}} < c'$ (Case I) or $q_l \ge c'$ (Case II). If w is small, then $c' = (1-w)q_u^{\text{ave}}$ will be close to q_u^{ave}, and therefore Case I will hold. However, conditions for Theorem 1 to apply will be found below for both Case I and Case II.

Case I: Let $q_l^{\text{ave}} < (1-w)q_u^{\text{ave}}$ (this corresponds to $(1-w)q_u^{\text{ave}}$ lying in the interval $[q_l^{\text{ave}}, q_u^{\text{ave}}]$. Then

$$g^3(a') = (1-w)^2 q_u^{\text{ave}} + w\left(\frac{NK}{\sqrt{\frac{p_{\max}((1-w)q_u^{\text{ave}} - q_{\min})}{(q_{\max} - q_{\min})}}} - \frac{dC}{M}\right).$$

Hence, the criterion $d' \ge a' > b' > c'$ of Theorem 1 ensuring the existence of chaos gives

$$(1-w)^2 q_u^{\text{ave}} + w\left(\frac{NK}{\sqrt{\frac{p_{\max}((1-w)q_u^{\text{ave}} - q_{\min})}{(q_{\max} - q_{\min})}}} - \frac{dC}{M}\right) \ge a'. \tag{27}$$

Case II: Alternatively, suppose $q_l^{\text{ave}} \ge (1-w)q_u^{\text{ave}}$. Then $g^3(a') = (1-w)^2 q_u^{\text{ave}} + wB$. Now the criterion $d' \ge a' > b' > c'$ of Theorem 1 ensuring the

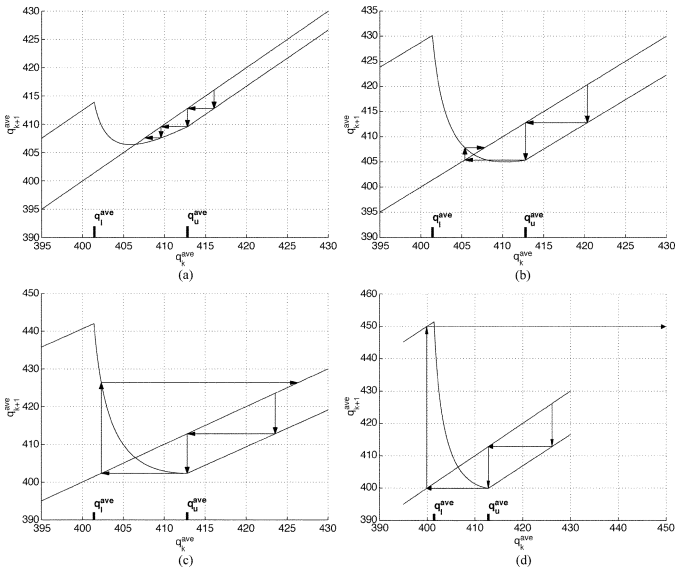

Fig. 9. First return map and period-three condition for (a) $w = 2^{-7}$, (b) for $w = 2^{-5.8}$, (c) for $w = 2^{-5.3}$, and (d) for $w = 2^{-5}$.

existence of chaos in this case gives $g^3(a') - a' \geq 0$, which reduces to following simple condition:

$$\begin{aligned}
&(1-w)^2 q_u^{\text{ave}} + wB \geq \frac{q_u^{\text{ave}}}{(1-w)} \\
&\Rightarrow (1-w)^3 q_u^{\text{ave}} + w(1-w)B \geq q_u^{\text{ave}} \\
&\Rightarrow w(1-w)B - \left(1-(1-w)^3\right) q_u^{\text{ave}} \geq 0 \\
&\Rightarrow w(1-w)B - (1-1+3w-3w^2+w^3) q_u^{\text{ave}} \geq 0 \\
&\Rightarrow q_u^{\text{ave}} \leq \frac{(1-w)B}{(3-3w+w^2)}. \qquad (28)
\end{aligned}$$

Summarizing, we have the following lemma. Here, "chaotic in the sense of Li and Yorke" means satisfying the conclusions of Theorem 1.

Theorem 2: The TCP-RED system given by (13) is chaotic in the sense of Li and Yorke if either (27) and $q_l^{\text{ave}} < (1-w) q_u^{\text{ave}}$ hold or (28) and $q_l^{\text{ave}} \geq (1-w) q_u^{\text{ave}}$ hold.

The progression of nonlinear instabilities toward period-three and chaos is illustrated in Fig. 9. It is shown that, as exponential averaging weight w is increased initially, the condition given by Case I holds, and for larger values of w the condition given by Case II is satisfied.

VIII. Simulation Results

In this section, we verify the existence of instabilities through the simulation results obtained using ns-2 simulator developed at the University of California at Berkeley and Lawrence Berkeley Laboratory (LBL). We demonstrate the parametric sensitivity of the RED to different system parameters, such as the number of connections and round-trip delays. The simulated network topology is as shown in Fig. 1. TCP connections are Reno connections. The propagation delays of the edge links that connect the sources to node $r1$ or node $r2$ to the destinations are uniform random variables selected from [10 ms, 35 ms]. The capacity of these edge links are set to 30 Mb/s.

In CA mode, a TCP Reno connection increases its congestion window size by one during an RTT if there is no packet drop. When it detects a packet loss, it sets the window size to roughly half of what it was at the time of detection. Thus, although it is a very robust mechanism, TCP leads to an oscillatory behavior due to this bandwidth estimation scheme in CA mode. Hence, in practice, with a small number of TCP connections, one would expect to see both the queue size and the average queue size fluctuate significantly. Such an oscillatory behavior will be mitigated with the increasing N, as demonstrated by Tinnakornsrisuphap and Makowski [36]. Thus, the oscillation in the average queue size per flow induced by this oscillatory behavior of TCP diminishes with the increasing number of flows. However, the oscillation induced by instability does not decrease with the number of flows, as should be apparent from Section III. We run the simulation with a reasonably large number of connections in order to reduce the fluctuation in the queue size due to the oscillatory behavior of TCP and set $N = 500$.

The capacity and delay of the bottleneck link are set to $C = 150$ Mb/s and 30 ms, respectively. Given these parameters, the average round-trip propagation delay (without any queuing delay) is approximately 150 ms with the standard deviation of 20.4 ms. The packet size for TCP connections is 500 bytes, and the buffer size at node $r1$ and $r2$ is 7500 packets. The threshold values $q_{\min}$ and $q_{\max}$ are set to 750 packets and 2250 packets, respectively, and $p_{\max}$ is set to 1/8. The simulation is run for 200 s. We plot the queue size and average queue size every 200 ms.

A. Effects of Exponential Averaging Weight

In this subsection, we vary the exponential averaging weight and study how it affects the queue behavior while other system and control parameters are fixed. Fig. 10 plots the evolution of the average queue size from $t = 100$ s to $t = 200$ s. In Fig. 10(a), the average queue size remains relatively stable around a single attractor or equilibrium, except for the fluctuations caused by the bandwidth estimation scheme of TCP, and there is no noticeable change in queue behavior. However, with increasing w_{red}, the average queue size begins to oscillate [Fig. 10(b)]. Finally, when w_{red} reaches some threshold value as predicted by our results, the average queue size shows no sign of any equilibrium point and constantly oscillates between the boundaries [Fig. 10(c) and (d)]. Such oscillations in queue size result in a frequently empty queue, as shown in Fig. 10(d) and, hence, lower throughput. The throughput decreases proportionally with the fraction of time the queue is empty, representing a waste of network resources. This example illustrates how one can expect very different behavior in queue sizes with a small change in the averaging weight w_{red}. Note that the exponential averaging weight is increased only by 25% from Fig. 10(b) to Fig. 10(d), clearly indicating a very sensitive nature of the RED to the parameter w_{red}.

We use this numerical example to further illustrate the relationship between our discrete-time model given by (4)–(6) and the simulation setup, which is first described in Section III. In this example, the average queue size when the system is stable,

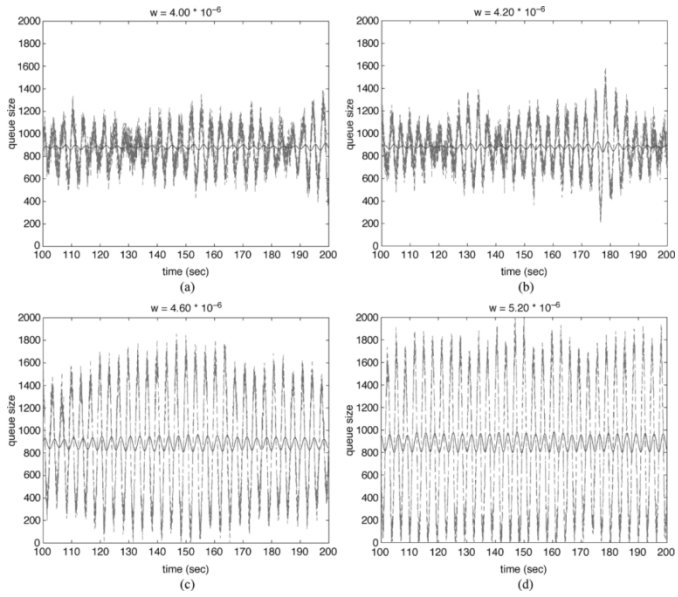

Fig. 10. Average queue sizes with various $w_{\rm red}$. (a) $w_{\rm red} = 4.0 \cdot 10^{-6}$. (b) $w_{\rm red} = 4.2 \cdot 10^{-6}$. (c) $w_{\rm red} = 4.6 \cdot 10^{-6}$, (d) $w_{\rm red} = 5.2 \cdot 10^{-6}$.

e.g., Fig. 10(a), is approximately 880 packets. We compute the value of K that is a solution to

$$\frac{NMK}{\sqrt{p^*}\left(d + \frac{q^* M}{C}\right)} = C$$

which is 1.35 ($\approx \sqrt{3/2} = 1.2247$). With $K = 1.35$ and $q^* = 880$ packets, (20) gives us $w_{\rm crit} = 0.0769$. Our simulation results indicate that the initial bifurcation takes place slightly after $w_{\rm red} = 4.2 \cdot 10^{-6}$ [Fig. 10(b)], and we approximate it to be $4.3 \cdot 10^{-6}$. Hence, due to the differences in the time scales of the discrete-time model and the ns-2 simulation as explained in Section III, the ratio of the averaging weight of our discrete-time model w to that of the RED gateway $w_{\rm red}$ in the neighborhood of the parameters used in this example is given by $(0.0769/(4.3 \cdot 10^{-6})) = 1.79 \cdot 10^4$.

We now compare the magnitude of oscillation in our simulation to the predicted magnitude from the bifurcation diagram shown in Fig. 11. Although it is difficult to perfectly estimate the oscillation magnitude due to the noise present in the simulation, the range of oscillation in Fig. 10(c) and (d) is approximately [840, 940] and [810, 990], respectively. The values of period-two fixed points are approximately 850 and 940 at $w = w_{\rm crit} \times ((4.6 \cdot 10^{-6})/(4.3 \cdot 10^{-6}))$ and 820 and 1000 at $w = w_{\rm crit} \times ((5.2 \cdot 10^{-6})/(4.3 \cdot 10^{-6}))$ from Fig. 11. Therefore, the magnitude of the observed oscillations in our simulation is reasonably close to the expected magnitude from the bifurcation diagram, in spite of the presence of noise due to a limited number of flows.

B. Effects of Round-Trip Propagation Delay d

This subsection demonstrates the sensitivity of the system stability to the round-trip delays of the connections. We use the same parameter values as in Section VIII-A except for the delay d, which is the bifurcation parameter in this subsection. The value of the averaging weight w in our discrete-time model is

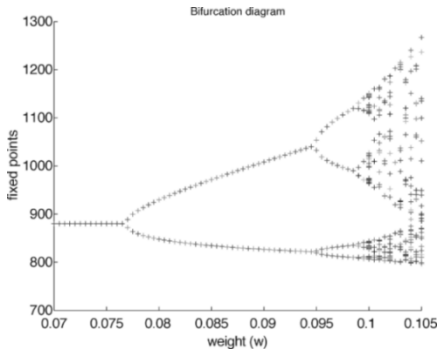

Fig. 11. Bifurcation of the example.

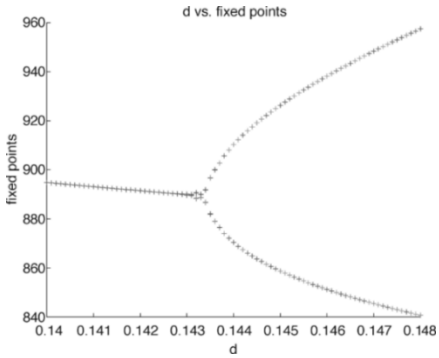

Fig. 12. Bifurcation diagram with varying d. ($w_{\rm red} = 0.085, N = 500$).

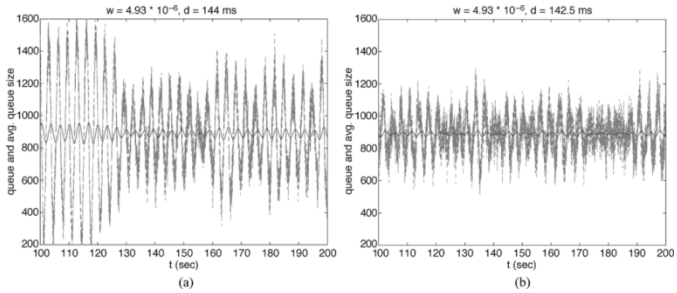

Fig. 13. Queue evolution with different delays. ($w_{\rm red} = 4.93 \cdot 10^{-6}$). (a) $d =$ 145 ms. (b) $d = 142.5$ ms.

set to 0.085, which corresponds to $w_{\rm red} = 4.7 \cdot 10^{-6}$ when $d =$ 150 ms. Fig. 12 plots the bifurcation diagram with varying d. It indicates that the system becomes unstable around $d = 143$ ms. For ns-2 simulation, we have fixed the averaging weight at $4.93 \cdot 10^{-6}$ and decreased the mean round-trip propagation delay of the connections to 145 and 142.5 ms by decreasing the one-way propagation delay of the bottleneck link accordingly. The value of $w_{\rm red}$ is increased slightly according to

$$\frac{150}{143} \times 4.7 \cdot 10^{-6} = 4.93 \times 10^{-6}$$

in order to compensate for a decrease in the round-trip delay and have the same averaging weight w in our discrete-time model. Note that the variation in queueing delay due to a change in d is much smaller than the change in d according to Fig. 12, and hence is negligible. Fig. 13 plots the queue evolution. These plots show that, as the delay is decreased to 145 ms and then to 142.5 ms, the system becomes stable as predicted by our model in Fig. 12.

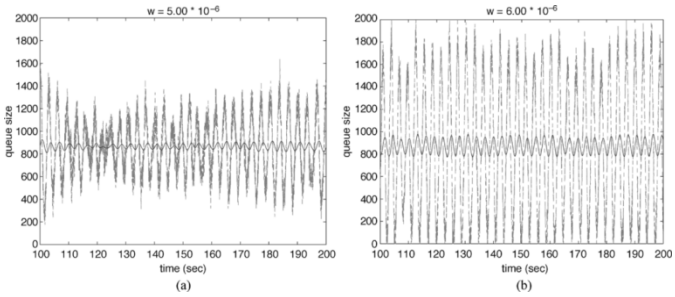

Fig. 14. Queue evolution with short-lived and long-lived connections. (a) $w_{\rm red} = 5.0 \cdot 10^{-6}$. (b) $w_{\rm red} = 6.0 \cdot 10^{-6}$.

C. Short-Lived Flows

In this subsection, we introduce short-lived flows and investigate how their presence affects system behavior. Here we reduce the number of long-lived TCP connections to 400 and introduce short-lived connections that arrive according to a Poisson process with $\lambda = 50$ connections/s. The duration of these short-lived connections is exponentially distributed with a mean of 2.0 s. This gives an average number of active connections of 500, as in the long-lived connection case, and approximately 80% of the traffic is generated by the long-lived connections.

As shown in Fig. 14, the queue behavior changes only slightly as the mixture of connections changes. Even with short-lived connections, the qualitative behavior of the queue size changes only marginally, and for sufficiently large averaging weight the queue size exhibits similar unstable behavior as in Fig. 10. The fact that the initial bifurcation point increases with short-lived flows can be explained as follows. The aggregate of the short-lived flows can be viewed as nonresponsive traffic since these connections may not live long enough to react to congestion indication before they complete the transfer and can be modeled as UDP traffic. It has been shown in [20] that the presence of UDP traffic enhances the system stability, which is intuitive because UDP traffic does not react to any congestion notification. A similar result is also proved in a rate control problem with a communication delay with nonresponsive flows [32].

IX. Conclusion

A nonlinear deterministic packet-level model of a TCP-RED gateway is developed and its dynamical behavior is analyzed. The analysis is the first to uncover rich nonlinear behaviors including multiple periodicities and chaotic behavior in TCP-RED models. The analysis agrees with previous results from linear modeling of congested networks and adds to these results by showing possible behaviors under wide system and control parameter variations. The first instability to occur was found to be a PDB, which contrasts with earlier reports of instability being caused by the discontinuity in the RED AQM law. This instability is followed by another bifurcation that is caused by the finite buffer size, namely a BCB. The discovery of this route to instability and associated erratic behavior is important for its own sake, but also for the investigation of robust control schemes to mitigate these instabilities. Numerical and ns-2 simulations were presented to provide further evidence and validation of the phenomena uncovered in this work. Further work of the authors [31] has uncovered a similar period-doubling instability in Kelly's optimal rate-allocation scheme [17].

References

[1] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamcal Systems*. New York: Springer-Verlag, 1996.

[2] S. Athuraliya, S. Low, V. H. Li, and Q. Yin, "REM: active queue management," *IEEE Network*, vol. 15, pp. 48–53, May/June 2001.

[3] F. Baccelli, D. R. McDonald, and J. Reynier, "A Mean-Field Model for Multiple TCP Connections Through a Buffer Implementing RED," INRIA, TR-4449, Apr. 2002.

[4] S. Banerjee, P. Ranjan, and C. Grebogi, "Bifurcations in two-dimensional piecewise smooth maps—theory and applications in switching circuits," *IEEE Trans. Circuits Syst. I*, vol. 47, pp. 633–643, May 2000.

[5] L. S. Brakmo and L. L. Peterson, "TCP Vegas: End to end congestion avoidance on a global internet," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1465–1480, Oct. 1995.

[6] V. Firoiu and M. Borden, "A study of active queue management for congestion control," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, 2000, pp. 1435–1444.

[7] S. Floyd, "TCP and explicit congestion notification," *ACM Computer Commun. Rev.*, vol. 24, pp. 10–23, Oct. 1994.

[8] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Networking*, vol. 1, pp. 397–413, Aug. 1993.

[9] R. J. Gibbens and F. Kelly. (1998, June) Resource Pricing and the Evolution of Congestion Control. [Online] http://www.statslab.cam.ac.uk/~frank

[10] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer-Verlag, 1983.

[11] J. P. Hespanha, S. Bohacek, K. Obraczka, and J. Lee, "Hybrid modeling of TCP congestion control," *Lecture Notes in Computer Science*, vol. 2034, pp. 291–304, 2001.

[12] C. V. Hollot, V. Misra, D. Towsley, and W. Gong, "A control theoretic analysis of RED," in *Proc. IEEE INFOCOM*, Anchorage, AK, 2001, pp. 1510–1519.

[13] ——, "On designing improved controllers for AQM routers supporting TCP flows," in *Proc. IEEE INFOCOM*, Anchorage, AK, 2001, pp. 1726–1734.

[14] E. A. Jackson, *Perspectives of Nonlinear Dynamics*. New York: Cambridge Univ. Press, 1991.

[15] V. Jacobson, "Congestion avoidance and control," *Computer Commun. Rev.*, vol. 18, no. 4, pp. 314–329, Aug. 1988.

[16] R. Johari and D. Tan, "End-to-end congestion control for the internet: delays and stability," *IEEE/ACM Trans. Networking*, vol. 9, pp. 818–832, Dec. 2001.

[17] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[18] S. Kunniyur and R. Srikant, "Analysis and design of adaptive virtual queue algorithm for active queue management," in *Proc. ACM SIGCOMM*, San Francisco, CA, 2001.

[19] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. Networking*, vol. 10, pp. 272–286, Apr. 2002.

[20] R. J. La, P. Ranjan, and E. H. Abed, "Nonlinearity of TCP and instability with RED," presented at the SPIE ITCom, Boston, MA, 2002.

[21] ——, "Analysis of Adaptive Random Early Detection (ARED)," presented at the 18th Int. Teletraffic Congr. (ITC), Berlin, Germany, Sept. 2003.

[22] T. Li and J. A. Yorke, "Period three implies chaos," *Amer. Math. Monthly*, vol. 82, no. 10, pp. 985–992, Dec. 1975.

[23] S. H. Low, F. Paganini, J. Wang, S. Adlakha, and J. C. Doyle, "Dynamics of TCP/RED and a scalable control," in *Proc. IEEE INFOCOM*, New York, 2002, pp. 239–248.

[24] S. J. Hogan, M. D. Bernardo, M. I. Feigin, and M. E. Homer, "Local analysis of c-bifurcations in n-dimensional piecewise smooth dynamical systems," *Chaos, Solitons Fractals*, vol. 10, no. 11, pp. 1881–1908, 1999.

[25] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The macroscopic behavior of the TCP congestion avoidance algorithm," *Computer Commun. Rev.*, vol. 27, no. 3, pp. 67–82, 1997.

[26] M. May, B. Lyles, J. Bolot, and C. Diot, "Reasons not to deploy RED," presented at the IFIP 5th Int. Workshop on Quality of Service (IWQOS '97), New York, 1997.

[27] H. E. Nusse and J. A. Yorke, "Border collision bifurcations including "period two to period three" for piecewise smooth maps," *Physica D*, vol. 57, no. 11, pp. 39–57, 1992.

[28] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Reno performance: a simple model and its empirical validation," *IEEE/ACM Trans. Networking*, vol. 8, pp. 133–145, Apr. 2000.
[29] P. Ranjan and E. H. Abed, "Nonlinear analysis and control of TCP-RED in a simple network model," presented at the Automatic Control Conf., Anchorage, AK, 2002.
[30] P. Ranjan, E. H. Abed, and R. J. La, "Nonlinear instabilities in TCP-RED," in *Proc. IEEE INFOCOM*, New York, 2002, pp. 249–258.
[31] ——, "Communication delay and instability in rate-controlled networks," in *Proc. 42nd IEEE Conf. Decision and Control*, Maui, HI, 2003, pp. 3689–3694.
[32] ——, (2003, Sept.) Trade-offs in rate control with communication delay. [Online] http://www.eng.umd.edu/~hyongla
[33] P. Ranjan, R. J. La, and E. H. Abed, "Controlling RED using washout filter controller," presented at the Automatic Control Conf., Boston, MA, 2004.
[34] A. N. Sharkovsky, S. F. Kolyada, A. G. Sivak, and V. V. Fedorenko, *Dynamics of One Dimensional Maps*. Norwell, MA: Kluwer, 1997.
[35] P. Tinnakornsrisuphap, R. J. La, and A. Makowski, "Characterization of general TCP Traffic under a large number of flows regime," presented at the 18th Int. Teletraffic Congr. (ITC), Berlin, Germany, Sept. 2003.
[36] P. Tinnakornsrisuphap and A. Makowski, "Limit behavior of ECN/RED gateways under a large number of TCP flows," in *Proc. IEEE INFOCOM*, Mar. 2003, pp. 873–883.
[37] R. Wolski and M. Swany. Network Weather Service. [Online]. Available: http://nws.cs.ucsb.edu/

Priya Ranjan received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, in 1997 and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1999 and 2003, respectively, all in electrical engineering.

He is currently a Research Associate with the Institute for Systems Research, University of Maryland, and a Research Scientist with Intelligent Automation, Inc., Rockville, MD. He has published in the areas of nonlinear dynamical systems, power electronics, modeling and analysis of logistic systems, and communication networking.

Eyad H. Abed (S'80–M'83–SM'93–F'02) received the S.B. degree from the Massachusetts Institute of Technology, Cambridge, in 1979, and the M.S. and Ph.D. degrees from the University of California, Berkeley, in 1981 and 1982, respectively, all in electrical engineering.

He has been with the Department of Electrical and Computer Engineering, University of Maryland, College Park, since 1983, where he is presently a Professor and Director of the Institute for Systems Research. His research interests center on control of nonlinear dynamical systems. Applications he has studied include electric power networks, aircraft engines, radar tracking systems, power electronics, supercavitating bodies, and computer/communication networks.

Dr. Abed was the recipient of the National Science Foundation Presidential Young Investigator Award, the O. Hugo Schuck Best Paper award from the American Automatic Control Council, the Senior Fulbright Scholar Award, the Alan Berman Research Publication Award from the Naval Research Laboratory, and several research and teaching awards from the University of Maryland. He currently serves on the Advisory Editorial Board of *Nonlinear Dynamics*.

Richard J. La received the B.S.E.E. degree from the University of Maryland, College Park, in 1994 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1997 and 2000, respectively.

From 2000 to 2001, he was a Senior Engineer with the Mathematics of Communication Networks Group, Motorola. Since August 2001, he has been on the faculty of the Electrical and Computer Engineering Department, University of Maryland. His research interests include resource allocation in communication networks and application of game theory.

Dr. La was the recipient of a National Science Foundation CAREER Award.