TECHNICAL RESEARCH REPORT

Limiting Model of ECN/RED under a Large Number of Heterogeneous TCP Flows

by Peerapol Tinnakornsrisuphap, Richard J. La

CSHCN TR 2003-7
(ISR TR 2003-13)

The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site http://www.isr.umd.edu/CSHCN/

Limiting Model of ECN/RED under a Large Number of Heterogeneous TCP Flows

Peerapol Tinnakornsrisuphap and Richard J. La
Department of Electrical and Computer Engineering and Institute for Systems Research
University of Maryland, College Park, MD, 20783
Email: {peerapol,hyongla}@eng.umd.edu

Abstract

Accurate modeling of a large number of heterogeneous TCP flows is important for the understanding and control of Internet traffic. Difficulties in deriving such models arise due to the interaction between different protocol layers and a large state space size. We introduce a stochastic model of a bottleneck ECN/RED gateway under a large number of competing heterogeneous TCP flows. Our main result shows that as the number of flows becomes large, the queue dynamics and the aggregate traffic are simplified and can be accurately described by simple statistical recursions. These recursions can be evaluated independently of the number of flows, and hence the resulting traffic model is scalable. We also present a simple analysis on the buffer utilization and window size at the steady-state. Simulation results supporting the theoretical findings are also presented.

I. Introduction

Due to the growing size and popularity of the Internet, Internet traffic modeling and control has become an important research area. Internet traffic consists of many heterogeneous traffic sources, the majority of which utilize Transmission Control Protocol (TCP) congestion control mechanism [3]. Characterization and modeling of TCP traffic yields an understanding of the interaction between the transport layer (TCP) and the network layer. Such interaction is well-understood in the context of a single long-lived TCP flow with a fixed loss probability [5]. However, the issues of modeling and understanding TCP traffic in a more realistic situation where there are many flows competing for bandwidth are much more challenging due to the following factors: (i) the explosion of state-space, (ii) Active Queue Management (AQM) schemes, (iii) session layer dynamics, and (iv) variable round-trip times (RTTs) of TCP flows.

When the number of TCP flows is large, straightforward modeling usually results in a model that is not scalable because of the explosion of the size of the state space required to model the state variables of the flows. Moreover, Internet TCP traffic is characterized by connections with diverse RTTs and their dynamic arrivals and departures. Little work has been done on modeling the interaction between an AQM mechanism and a large number of general TCP flows with variable delays. Some initial investigation of the role of heterogeneous RTTs in such an interaction is presented in [4]. However, the study is limited to a small number of TCP flows, *e.g.,* less than a hundred flows. Additional modeling difficulties arise from the fact that the feedback information (*i.e.,* marks on packets) from the AQM mechanism to TCP flows arrives at different rate depending on the RTTs of the connections. These obstacles present a considerable difficulty in deriving a scalable model that can capture the important aspect of Internet traffic dynamics and yield insights into how to control it. In order to deal with such difficulties typically some ad-hoc assumptions are made to simplify the model. As a result, the models become accurate only in certain regimes. The shortcomings of these models suggest a need for a *unified* model that is accurate in *all* regimes, instead of being restricted to a specific regime.

In this paper, we present a novel approach to modeling an ECN/RED bottleneck with a large number of TCP flows [1], [2]. Our model incorporates not only the interaction of congestion control mechanism of TCP with ECN/RED mechanism, but also session dynamics and variable RTTs of the flows. It builds upon the approach used in [7] and [8]. Such "macroscale" modeling of aggregate TCP flows can be developed

by systematically applying the limit theorems to derive a limiting traffic model when the number of TCP flows is large. When appropriate limit theorems are applied, typically model simplification occurs without having to rely on ad-hoc assumptions. Based on our model we show that the queue size per session and the workload per session brought in during a RTT converge to a deterministic process as the number of flows increases. We also demonstrate that the flows become asymptotically independent, which indicates that the RED mechanism indeed helps break the synchronization among the flows suffered by drop-tail gateways. In addition, based on our results we present a simple analysis on the buffer utilization and window size at the steady-state, which suggests that only the mean RTT affects the mean queue size at the steady-state.

This paper is organized as follows. Section II introduces the model. Then in Section III, the main theorem of the paper is presented. An analysis on the buffer utilization at the steady-state is presented in Section IV. We discuss the results in Section V along with conclusions and suggestions for future work.

II. The Model

In our model, we have three layers of dynamics, namely network, transport, and session layers, which interact with each other through mechanisms that will be specified shortly. At the lowest level, the network is simplified to be a single bottleneck router with an ECN/RED marking mechanism controlling the congestion level. The traffic injected into the network is controlled by TCP congestion control mechanism in the transport layer, which reacts to the marks from the network. Each TCP connection is initiated by a session. A session can be either active or idle. If a session is busy, a file or an object is transferred through a TCP connection. A busy period of a session lasts until it no longer has any more data to transfer, at which time it goes idle. The duration of an idle period is random and represents the idle time between consecutive file transmissions. When a new file/object to be transferred arrives, the session becomes active again and sets up a new TCP connection. We now give detailed descriptions of the model for each layer and the interaction of these three layers.

Let $\mathcal{N} = \{1, \cdots, N\}$ be the set of sessions that share a bottleneck RED gateway. Time is assumed to be slotted into contiguous timeslots, where a timeslot is the greatest common divider of the RTTs of TCP flows. We write $X^{(N)}$ to indicate the explicit dependence of the quantity X on the number N of sessions. Equivalence in law or in distribution between random variables (rvs) is denoted by $=_{st}$. The indicator function of an event A is given by $\mathbf{1}[A]$, and we use $\stackrel{P}{\to}_n$ (resp. $\Longrightarrow_n$) to denote convergence in probability (resp. weak convergence or convergence in distribution) with n going to infinity. An expectation of a rv X with a distribution function F is given by either $\mathbf{E}[X]$ or $\mathbf{E}[F]$.

A. Heterogeneous Round-trip Times

We assume that the RTTs of TCP connections are integer multiples of timeslots, and any congestion-control actions by TCP flows, *i.e.*, additive increase and multiplicative decrease, occur at the end of the round-trip. The RTT of flow i at time t is denoted by $d_i^{(N)}(t) \in \mathcal{H} := \{2, 3, \cdots, D_{max}\}$ [1] We use $\beta_i^{(N)}(t+1)$ to denote the number of timeslots since the last action by an *active* flow i. Then, $\beta_i^{(N)}(t)$ evolves according to

$$\beta_i^{(N)}(t+1) = \left(1 + \beta_i^{(N)}(t)\mathbf{1}\left[\beta_i^{(N)}(t) < d_i^{(N)}(t)\right]\right) \times \mathbf{1}\left[X_i^{(N)}(t) > 0\right], \tag{1}$$

where $X_i^{(N)}(t)$ is the remaining workload (in packets) of connection i at the beginning of timeslot $[t, t+1)$. The rv $X_i^{(N)}(t) > 0$ only if connection i is active in timeslot $[t, t+1)$. Hence, the last indicator function is one only if the connection is active. This will be explained further in the next subsection.

We use the following mapping $G_{i,t} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ to simplify our notation later.

$$G_{i,s}(a, b) = a \cdot \mathbf{1}\left[\beta_i^{(N)}(s) < d_i^{(N)}(s)\right] + b \cdot \mathbf{1}\left[\beta_i^{(N)}(s) \geq d_i^{(N)}(s)\right]. \tag{2}$$

[1] While one can modify (39) in the proof to incorporate $1 \in \mathcal{H}$ in the model, the absence of $1 \in \mathcal{H}$ does not cause any loss in the generality of the model.

If we let $Y_i^{(N)}(t+1) = G_{i,t+1}(Y_i^{(N)}(t), Y_i^{new})$, then the values of $Y_i^{(N)}(t+1)$ will be updated to Y_i^{new} only at the end of the round-trip, *i.e.,* $\beta_i^{(N)}(t+1) \geq d_i^{(N)}(t+1)$. Otherwise, $Y_i^{(N)}(t+1) = Y_i^{(N)}(t)$ since no action will be taken before the round-trip.

B. Session Dynamics

Each session $i \in \mathcal{N}$ is either active or idle. An idle session at the beginning of timeslot $[t, t+1)$ does not have any packets to transmit in the timeslot. An idle session in timeslot $[t, t+1)$ becomes active at the beginning of timeslot $[t+1, t+2)$ with probability $P_{ar} \in (0,1)$ independently of the past. In other words, the duration of an idle period is geometrically distributed with parameter P_{ar} and has a mean of $1/P_{ar}$. This attempts to capture the dynamics of connection arrivals, where the interarrival times are reported to be exponentially distributed [6]. [2] Let $\{U_i(t), i \in \mathcal{N}; t = 0, 1, \cdots\}$ be a collection of i.i.d. rvs uniformly distributed on [0, 1] and $\mathbf{1}\left[U_i(t+1) < P_{ar}\right]$ be the indicator function of the event that a new file/object arrives in the timeslot $[t+1, t+2)$ for an idle session i.

Let $\{F_i(t), i \in \mathcal{N}; t = 0, 1, \cdots\}$ be a collection of i.i.d. non-negative integer-valued rvs with a general distribution function F. The workload of a connection of session i that becomes active at the beginning of timeslot $[t, t+1)$ is given by $F_i(t)$. This workload represents the *total* amount of workload a TCP connection brings in before it is torn down rather than workload brought in by an object or a file. In other words, if the same TCP connection is used to transfer more than one object while it is alive, $F_i(t)$ represents the total amount of workload brought in by all objects during the active period. The evolution of $X_i(t)$, which denotes the remaining workload, is given by the following:

$$\begin{aligned} X_i^{(N)}(t+1) = \mathbf{1}\left[X_i^{(N)}(t) > 0\right]\left(X_i^{(N)}(t) - A_i^{(N)}(t)\right) \\ + \mathbf{1}\left[X_i^{(N)}(t) = 0\right]\mathbf{1}\left[U_i(t+1) < P_{ar}\right]F_i(t+1)\,, \end{aligned} \tag{3}$$

where $A_i^{(N)}(t)$ denotes the number of packets injected into the network by connection i at the beginning of timeslot $[t, t+1)$. This will be explained in the following subsection.

When a new connection arrives, its RTT is randomly selected, and the RTT of session i at timeslot $[t+1, t+2)$ is given by

$$d_i^{(N)}(t+1) = d_i^{(N)}(t)\mathbf{1}\left[X_i^{(N)}(t) > 0\right] + \mathbf{1}\left[X_i^{(N)}(t) = 0\right]\mathbf{1}\left[U_i(t+1) < P_{ar}\right]H_i(t+1)\,, \tag{4}$$

where $H_i(t+1)$'s are i.i.d. rvs that take values in $\mathcal{H}$ and determine the new RTTs of newly arrived connections. The bound on the maximum RTT does not constrain the model because actual TCP flows also cannot have larger RTTs than the timeout value.

C. TCP Dynamics

For each $i \in \mathcal{N}$, let $W_i^{(N)}(t)$ be an integer-valued rv that encodes the congestion window size (in packets) at the beginning of timeslot $[t, t+1)$. We assume that the range of rv $W_i^{(N)}(t)$ is $\{0, 1, \cdots, W_{\max}\}$, where $W_{\max}$ is a finite integer representing the receiver advertised window size of the TCP connection. We assume that the congestion window size of an idle session is zero. When an idle session becomes active at the beginning of timeslot $[t, t+1)$, the congestion window size of TCP connection is set to one at the end of the first round-trip, *i.e.,* at timeslot $[t + d_i^{(N)}(t), t + d_i^{(N)}(t) + 1)$, allowing a transmission of one packet. This models one RTT for three-way handshake. Here we describe how the congestion window sizes of active connections evolve.

Each TCP source transmits as many of the remaining data packets as allowed by its congestion window only at the end of the round-trip. We simplify the packet transmission in the round-trip so that the packets

[2] Recall that one can approximate an exponential rv X with parameter α with $\lceil X \rceil$, which is a geometric rv with parameter $p = 1 - e^{-\alpha}$.

from a connection all arrive only in a single timeslot, rather than being spread out throughout a round-trip. Such simplification can be justified by the following:

(i) In the Internet, most of the packet arrivals at a bottleneck are usually compressed together due to the "ACK compression" phenomenon [9], which leads to bursty arrivals at the bottlenecks. Hence, modeling the packet arrivals over a RTT as a batch arrival in a single timeslot tends to be more accurate than modeling them as smooth arrivals throughout the RTT.

(ii) Aggregating a round-trip worth of packet arrivals into a single timeslot will result in burstier traffic from each flow. This will cause queue dynamics to fluctuate more than having a smooth arrival pattern. Therefore, the queue fluctuation in this model will provide an upper bound to the actual queue with smoother packet arrival patterns.

(iii) The information used for control action at the RED gateways is the average queue size. With the averaging mechanism with long memory as in RED, the difference in the queue size due to our bursty packet arrivals will be smoothed out by the averaging mechanism of RED.

Suppose that connection i has $X_i^{(N)}(t)$ remaining packets (or workload) waiting to be transmitted at the beginning of timeslot $[t, t+1)$.[3] The number of packets connection i transmits at the beginning of timeslot $[t, t+1)$, denoted by $A_i^{(N)}(t)$, is given by

$$A_i^{(N)}(t) = \min\left(W_i^{(N)}(t), X_i^{(N)}(t)\right) \mathbf{1}\left[\beta_i^{(N)}(t) \geq d_i^{(N)}(t)\right]. \tag{5}$$

Note from (5) that a connection transmits once per RTT.

The congestion control mechanism of TCP operates in two different modes: slow start (SS) and congestion avoidance (CA). A new TCP connection starts in SS. In SS, the congestion window size is doubled every RTT until one or more packets are marked. If a mark is received, then the congestion window size is halved and TCP switches to CA. The congestion window size is limited by the receiver advertised window size $W_{\max}$. Hence, the evolution of the congestion window of connection i in SS can be written as

$$\begin{aligned} W_{i,SS}^{(N)}(t+1) &= G_{i,t+1}[W_i^{(N)}(t), \min\left(2W_i^{(N)}(t) \vee 1, W_{\max}\right) M_i^{(N)}(t+1) \\ &\quad + \min\left(\lceil \frac{W_i^{(N)}(t)}{2} \rceil, W_{\max}\right)(1 - M_i^{(N)}(t+1))], \end{aligned} \tag{6}$$

where $a \vee b = \max(a, b)$ and $M_i^{(N)}(t+1)$ is an indicator function of the event that no marks have been received in the round-trip preceding the timeslot $[t, t+1)$, *i.e.*, $M_i^{(N)}(t+1) = 1$ when no packet from Session i is marked in the previous round-trip and $M_i^{(N)}(t+1) = 0$ when at least one packet is marked. The marking mechanism will be explained in more detail in Subsection II-D. From the definition of mapping in (2), the window size is updated only once per RTT.

In CA, the congestion window size in the next timeslot is increased by 1 if no marks are received in round-trip preceding the timeslot $[t, t+1)$, and if one or more packets are marked the congestion window is reduced by half. The congestion window size in CA can be described by the following:

$$\begin{aligned} W_{i,CA}^{(N)}(t+1) &= G_{i,t+1}[W_i^{(N)}(t), \min(W_i^{(N)}(t) + 1, W_{\max}) M_i^{(N)}(t+1) \\ &\quad + \min\left(\lceil \frac{W_i^{(N)}(t)}{2} \rceil, W_{\max}\right)(1 - M_i^{(N)}(t+1))]. \end{aligned} \tag{7}$$

We use $\{0,1\}$-valued rvs $\{S_i^{(N)}(t), i \in \mathcal{N}\}$ to encode the state of TCP connections. We interpret $S_i^{(N)}(t) = 0$ (resp. $S_i^{(N)}(t) = 1$) as connection i being in CA (resp. in SS) at the beginning of the timeslot

[3] We refer to a TCP connection of an active Session i by connection i when there is no confusion.

$[t, t+1)$. Therefore, the complete recursion of the congestion window size can be written as

$$W_i^{(N)}(t+1) = \mathbf{1}\left[X_i^{(N)}(t) - A_i^{(N)}(t) > 0\right] \times [S_i^{(N)}(t) W_{i,SS}(t+1) + (1 - S_i^{(N)}(t)) W_{i,CA}(t+1)] , \tag{8}$$

where the first indicator function is used to reset the congestion window size to zero when Session i runs out of data to transmit and returns to its idle state.

Finally, the evolution of $S_i^{(N)}(t)$ is given by

$$S_i^{(N)}(t+1) = \mathbf{1}\left[X_i^{(N)}(t) - A_i^{(N)}(t) \leq 0\right] + \mathbf{1}\left[X_i^{(N)}(t) - A_i^{(N)}(t) > 0\right] S_i^{(N)}(t) M_i^{(N)}(t+1). \tag{9}$$

This equation can be interpreted as follows. Connection i is in SS in timeslot $[t+1, t+2)$ if either (1) there is no packet left to transmit (so the connection resets) at the beginning of the timeslot or (2) the connection was active and in SS in timeslot $[t, t+1)$ and received no mark in the timeslot. From (9), we assume that a new TCP connection in SS is ready to be set up one timeslot after the previous connection is torn down after finishing its workload, and the new TCP connection becomes active when a new file/object arrives initiating three-way handshake. We also assume that the slow start/congestion avoidance state is updated in the next timeslot following transmission. However, the window size is updated one RTT after transmission using the appropriate SS/CA state as in the correct operation of TCP.

D. Network Dynamics

In this subsection we explain how packets are marked to provide the congestion notification to the active TCP connections. The capacity of the bottleneck link is NC packets/slot for some positive constant C. The buffer size is assumed to be infinite so that no packets are dropped due to buffer overflow. Thus, congestion control is achieved solely through the random marking algorithm of the RED gateway.

Let $Q^{(N)}(t)$ denote the number of packets queued in the buffer at the beginning of timeslot $[t, t+1)$. Connection i injects $A_i^{(N)}(t)$ packets into the network, and they are put in the buffer at the beginning of timeslot $[t, t+1)$. Let the rv

$$A^{(N)}(t) := \sum_{i=1}^{N} A_i^{(N)}(t) \tag{10}$$

denote the aggregate number of packets offered to the network by the N sessions at the beginning of timeslot $[t, t+1)$. Hence, $Q^{(N)}(t) + A^{(N)}(t)$ packets are available for transmission during that timeslot. Since the bottleneck link has a capacity of NC packets/timeslot, $\left[Q^{(N)}(t) + A^{(N)}(t) - NC\right]^+$ packets will not be served during timeslot $[t, t+1)$, and will remain in the buffer. Hence, their transmission is deferred to subsequent timeslots. The number of packets in the buffer at the beginning of timeslot $[t+1, t+2)$, $Q^{(N)}(t+1)$, is therefore given by

$$Q^{(N)}(t+1) = \left[Q^{(N)}(t) - NC + A^{(N)}(t)\right]^+ . \tag{11}$$

And the average queue $\hat{Q}^{(N)}(t)$ is given by

$$\hat{Q}^{(N)}(t+1) = (1-\alpha)\hat{Q}^{(N)}(t) + \alpha Q^{(N)}(t+1), \tag{12}$$

where $0 < \alpha \leq 1$ is the parameter of the exponential averaging mechanism.

Each incoming packet into the router in timeslot $[t, t+1)$ is marked with a probability $f^{(N)}\left(\hat{Q}^{(N)}(t)\right)$, depending on the average queue length at the beginning of the timeslot $[t, t+1)$. We represent this possibility

by the $\{0,1\}$-valued rvs $M_{i,j}^{(N)}(t+1)$ ($j=1,...,A_i^{(N)}(t)$) with the interpretation that $M_{i,j}^{(N)}(t+1)=0$ (resp. $M_{i,j}^{(N)}(t+1)=1$) if the jth packet from source i is marked (resp. not marked) in the RED buffer. To do so we introduce a collection of i.i.d. $[0,1]$-*uniform* rvs $\{V_{i,j}(t+1),\ i,j=1,\cdots;\ t=0,1,\cdots\}$ that are assumed to be independent of other rvs. The process by which packets are marked is as follows. For each $i\in\mathcal{N}$ and $j=1,2,\ldots$, we define the marking rvs

$$M_{i,j}^{(N)}(t+1)=\mathbf{1}\left[V_{i,j}(t+1)>f^{(N)}(\hat{Q}^{(N)}(t))\right],$$

so that the rv $M_{i,j}^{(N)}(t+1)$ is the indicator function of the event that the jth packet from source i is *not* marked in timeslot $[t,t+1)$. The indicator function of the event that no packets from connection i in timeslot $[t,t+1)$ are marked can now be written as $\prod_{j=1}^{A_i^{(N)}(t)}M_{i,j}^{(N)}(t+1)$. This information will be available to the TCP sender in the next timeslot. However, this information is used only one RTT later to update the congestion window size, and $M_i^{(N)}(t+1)$ evolves according to

$$M_i^{(N)}(t+1)=G_{i,t}(M_i^{(N)}(t),\ \prod_{j=1}^{A_i^{(N)}(t)}M_{i,j}^{(N)}(t+1)),\tag{13}$$

where we define $\prod_{j=1}^{0}M_{i,j}^{(N)}(t+1)=1$. Notice that we use time parameter t for the mapping G to delay the change in the value of $M_i^{(N)}$ by one timeslot, therefore, (6) and (7) evolve based on the markings in the previous round-trip as they should.

III. The Asymptotics

The main result of the paper consists of the asymptotics for the normalized buffer content as the number of sessions becomes large. This result is discussed under the following assumptions (A1)-(A2):

(A1) There exists a continuous function $f:\mathbb{R}_+\to[0,1]$ such that for each $N=1,2,\ldots$,

$$f^{(N)}(x)=f(N^{-1}x),\quad x\geq 0;$$

(A2) For each $N=1,2,\ldots$, the dynamics (3), (8), (9) and (11) start with the initial conditions

$$Q^{(N)}(0)=W_i^{(N)}(0)=\beta_i^{(N)}(0)=d_i^{(N)}(0)=0;\text{and }S_i^{(N)}(0)=M_i^{(N)}(0)=1;\quad i=1,\ldots,N.$$

We denote the vector of state variables for user i at timeslot $[t,t+1)$,

$$\mathbf{Y}_i^{(N)}(t):=(W_i^{(N)}(t),X_i^{(N)}(t),S_i^{(N)}(t),d_i^{(N)}(t),\beta_i^{(N)}(t),M_i^{(N)}(t)).\tag{14}$$

Assumption (A1) is a structural condition while (A2) is made essentially for technical convenience as it implies that for each N and all $t=0,1,\cdots$, the random vectors $\mathbf{Y}_1^{(N)}(t),\cdots,\mathbf{Y}_N^{(N)}(t)$ are *exchangeable*. Assumption (A2) can be omitted but at the expense of a more cumbersome discussion.

Theorem 1: Assume that (A1)-(A2) hold. Then, for each $N=1,2,\ldots$ and $t=0,1,\ldots$, there exists a (non-random) constant $q(t)$ and random vector $\mathbf{Y}(t)=(W(t),X(t),S(t),d(t),\beta(t),M(t))$ such that the following holds:

(i) *The following convergences take places:*

$$\frac{Q^{(N)}(t)}{N}\ \xrightarrow{P}{}_N\ q(t)\quad\textit{and}\quad\frac{\hat{Q}^{(N)}(t)}{N}\xrightarrow{P}\hat{q}(t)\tag{15}$$

$$\mathbf{Y}_1^{(N)}(t)\ \Rightarrow_N\ \mathbf{Y}(t)\tag{16}$$

(ii) *For any bounded function* $g : \mathbb{Z}_+^6 \to \mathbb{R}$

$$\frac{1}{N}\sum_{i=1}^{N} g\left(\mathbf{Y}(t)\right) \overset{P}{\to}_N \mathbf{E}\left[g\left(\mathbf{Y}(t)\right)\right], \tag{17}$$

$$\textit{Also, } \frac{1}{N}\sum_{i=1}^{N} A_i^{(N)}(t) \overset{P}{\to}_N \mathbf{E}\left[\min\left(W(t), X(t)\right)\mathbf{1}\left[\beta(t) \geq d(t)\right]\right]. \tag{18}$$

Moreover, if the workload distribution F *has a finite second moment, then*

$$\frac{1}{N}\sum_{i=1}^{N} X_i^{(N)}(t) \overset{P}{\to}_N \mathbf{E}\left[X(t)\right] \tag{19}$$

(iii) *For any integer* $I = 1, 2, \ldots,$ *the random vector* $\{\mathbf{Y}_i^{(N)}(t), i = 1, \ldots, I\}$ *becomes asymptotically independent as N becomes large, with*

$$\begin{aligned} &\lim_{N\to\infty} \mathbf{P}[\mathbf{Y}_i^{(N)}(t) = (w_i, x_i, s_i, d_i, \beta_i, m_i),\ i = 1, \cdots, I] \\ &= \prod_{i=1}^{I} \mathbf{P}\left[\mathbf{Y}(t) = (w_i, x_i, s_i, d_i, \beta_i, m_i)\right] \end{aligned} \tag{20}$$

for any $w_i, x_i, s_i, d_i, \beta_i, m_i, i = 1, \ldots, I$ *in* $\mathbb{Z}_+$.

In addition, with initial conditions $q(0) = W(0) = X(0) = d(0) = \beta(0) = 0, S(0) = M(0) = 1$*, it holds that*

$$q(t+1) = \left(q(t) - C + \mathbf{E}\left[A(t)\right]\right)^+ \tag{21}$$

$$\hat{q}(t+1) = (1-\alpha)\hat{q}(t) + \alpha q(t+1) \tag{22}$$

where $A(t) = \min\left(W(t), X(t)\right)\mathbf{1}\left[\beta(t) \geq d(t)\right]$*. Further, the recurrence*

$$\begin{aligned} \mathbf{Y}(t+1) &= \left(W(t+1), X(t+1), S(t+1), d(t+1), \beta(t+1), M(t+1)\right) \\ &=_{st} P(\mathbf{Y}(t)) := \left(P_1(\mathbf{Y}(t)), P_2(\mathbf{Y}(t)), \ldots, P_6(\mathbf{Y}(t))\right) \end{aligned} \tag{23}$$

holds in law, where

$$\begin{aligned} P_1(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) - A(t) > 0\right]\left(S(t)W_{SS}(t+1) + (1 - S(t))W_{CA}(t+1)\right) \\ P_2(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) > 0\right]\left(X(t) - A(t)\right)\mathbf{1}\left[X(t) = 0\right]\mathbf{1}\left[U(t+1) < P_{ar}\right]F(t+1) \\ P_3(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) - A(t) \leq 0\right] + \mathbf{1}\left[X(t) - A(t) > 0\right]S(t)M(t+1) \\ P_4(\mathbf{Y}(t)) &= d(t)\mathbf{1}\left[X(t) > 0\right] + \mathbf{1}\left[X(t) = 0\right]\mathbf{1}\left[U(t+1) < P_{ar}\right]H(t+1) \\ P_5(\mathbf{Y}(t)) &= \left(1 + \beta(t)\mathbf{1}\left[\beta(t) < d(t)\right]\right)\mathbf{1}\left[X(t) > 0\right] \\ G_s(a,b) &= a\mathbf{1}\left[\beta(s) < d(s)\right] + b\mathbf{1}\left[\beta(s) \geq d(s)\right] \\ P_6(\mathbf{Y}(t)) &= G_t(M(t), \mathbf{1}\left[V(t+1) \leq (1 - f(\hat{q}(t)))^{A(t)}\right]) \end{aligned} \tag{24}$$

$$\begin{aligned} W_{SS}(t+1) &= G_{t+1}\left(W(t), \min\left(2W(t) \vee 1, W_{\max}\right)M(t+1) + \lceil\frac{W(t)}{2}\rceil(1 - M(t+1))\right) \\ W_{CA}(t+1) &= G_{t+1}\left(W(t), \min\left(W(t)+1, W_{\max}\right)M(t+1) + \lceil\frac{W(t)}{2}\rceil(1 - M(t+1))\right) \end{aligned} \tag{25}$$

for i.i.d. $[0,1]$*-uniform rvs* $\{U(t+1), V(t+1);\ t = 0, 1, \ldots\}$.

Proof: The proof is given in Appendix I. ∎

IV. STEADY-STATE REGIME

We now turn our attention to the steady state regime of the limiting recursion (21)-(23), more specifically the calculation of the limiting queue and average window size in statistical equilibrium, *i.e.,* large t asymptotics.

Throughout this section we make the following assumptions.

(A3) The marking function $f : \mathbb{R} \to [0, 1]$ is monotonically increasing with

$$f(0) = 0 \quad \text{and} \quad \lim_{x\to\infty} f(x) = 1;$$

(A4) The sequence $\{(q(t), \mathbf{Y}(t)),\ t = 0, 1, \ldots\}$ admits a steady state in the sense that

$$(q(t), \mathbf{Y}(t)) \Rightarrow_t (q^\star, \mathbf{Y}^\star)$$

for some rvs $(q^\star, \mathbf{Y}^\star)$ where $q^\star$ is a constant and $\mathbf{Y}^\star = (W^\star, X^\star, S^\star, d^\star, \beta^\star, M^\star)$ is a $\{1, \ldots, W_{\max}\} \times \mathbb{Z}_+ \times \{0, 1\} \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0, 1\}$-valued rv.

(A5) Let F_{ar} be a rv with the distribution F, which represents the initial workload size of a new connection. We assume that $\mathbf{E}[F_{ar}] >> \mathbf{E}[W^\star]$. Since most of of the Internet traffic is generated by fewer number of long-lived connections, this is a reasonable assumption.

(A6) We assume that when an active connection finishes its last transmission, it waits an additional RTT before resetting its window size to zero. [4]

It is easily seen that Assumption (A4) immediately implies

$$\hat{q}(t) \xrightarrow{P}_t \hat{q}^\star = q^\star$$

for some constant $\hat{q}^\star$. And the steady-state marking probability is $f(\hat{q}^\star) = f(q^\star)$.

We wish to find the steady-state queue level $q^\star$ as a fixed-point solution to

$$\begin{aligned} q^\star &= (q^\star - C + \mathbf{E}[A])^+ \\ &= (q^\star - C + \mathbf{E}[\min(W^\star, X^\star)\mathbf{1}[\beta^\star \geq d^\star]])^+ \end{aligned} \tag{26}$$

Lemma 1: Assuming (A1)-(A6), the rvs $\min(W^\star, X^\star)$ *and* $\mathbf{1}[\beta^\star \geq d^\star]$ *are independent conditional on the event that the connection is active.*

Proof: First, the marking probability is fixed by (A4) and each of the session at the steady-state is independent from each other. Recall that the size of the workload as the connection is initiated and the round-trip delay are independent. We notice that conditioning on the event that the connection is active, $\mathbf{P}[\min(W^\star, X^\star) = w | d^\star = d, \text{active}] = \mathbf{P}[\min(W^\star, X^\star) = w | \text{active}]$. This is because the distribution of rv $\min(W^\star, X^\star)$ during the transmission time is independent of $d^\star$ (this can be proven by induction as a consequence of the fixed marking probability (A4)) and rv $\min(W^\star, X^\star)$ in between the transmission time has the same value as $\min(W^\star, X^\star)$ at the latest transmission time.

This also implies

$$\mathbf{P}[d^\star = d | \min(W^\star, X^\star) = w, \text{active}] = \mathbf{P}[d^\star = d | \text{active}]\ .^{5}$$

Furthermore,

$$\mathbf{P}[\beta^\star \geq d^\star | \min(W^\star, X^\star) = w, \text{active}, d^\star = d] = \mathbf{P}[\beta^\star \geq d^\star | \text{active}, d^\star = d] \tag{27}$$

[4] This will have only a marginal effect under the assumption $\mathbf{E}[F_{ar}] >> \mathbf{E}[W^\star]$.

[5] This can be viewed as a consequence of the fact that the initial workload rv and the RTT rv are independent.

because once the connection is active and the round-trip time is d, the probability of the counter $\beta^\star$ greater or equal than d does not depend on $\min(W^\star, X^\star)$. Finally,

$$\begin{aligned}
& \mathbf{P}\left[\beta^\star \geq d^\star | \min(W^\star, X^\star) = w, \text{active}\right] \\
= & \sum_{d \in \mathcal{H}} \mathbf{P}\left[\beta^\star \geq d^\star | \min(W^\star, X^\star) = w, \text{active}, d^\star = d\right] \mathbf{P}\left[d^\star = d | \min(W^\star, X^\star) = w, \text{active}\right] \\
= & \sum_{d \in \mathcal{H}} \mathbf{P}\left[\beta^\star \geq d^\star | \text{active}, d^\star = d\right] \mathbf{P}\left[d^\star = d | \min(W^\star, X^\star) = w, \text{active}\right] \\
= & \sum_{d \in \mathcal{H}} \mathbf{P}\left[\beta^\star \geq d^\star | \text{active}, d^\star = d\right] \mathbf{P}\left[d^\star = d | \text{active}\right] \\
= & \mathbf{P}\left[\beta^\star \geq d^\star | \text{active}\right]
\end{aligned}$$

■

The window size and the workload are both zero when the session is inactive. Hence, if we denote by $\mathbf{P}\left[\text{active}\right]$ the probability that a session is active at the steady-state, then

$$\begin{aligned}
\mathbf{E}\left[A^\star\right] &= \mathbf{P}\left[\text{active}\right] \mathbf{E}\left[\min(W^\star, X^\star) \mathbf{1}\left[\beta^\star \geq d^\star\right] | \text{active}\right] \\
&= \mathbf{P}\left[\text{active}\right] \mathbf{E}\left[\min(W^\star, X^\star) | \text{active}\right] \mathbf{P}\left[\beta^\star \geq d^\star | \text{active}\right]
\end{aligned} \tag{28}$$

where the last equality follows from Lemma 1.

First, consider

$$\mathbf{P}\left[\beta^\star \geq d^\star | \text{active}\right] = \sum_{d_i} \mathbf{P}\left[\beta^\star \geq d_i | \text{active}, d^\star = d_i\right] \mathbf{P}\left[d^\star = d_i | \text{active}\right]$$

From (A6), conditioning on the event that $d^\star = d$, it is easy to see that

$$\mathbf{P}\left[d^\star = d_i | \text{active}\right] = \frac{d_i \mathbf{P}\left[H = d_i\right]}{\sum_{d_j \in \mathcal{H}} d_j \mathbf{P}\left[H = d_j\right]}. \tag{29}$$

Under assumption (A5), since a connection typically lasts many RTTs, we have

$$\mathbf{P}\left[\beta^\star \geq d_i | \text{active}, d^\star = d_i\right] \approx \frac{1}{d_i}. \tag{30}$$

Therefore,

$$\begin{aligned}
\mathbf{P}\left[\beta^\star \geq d^\star | \text{active}\right] &\approx \sum_{d_i \in \mathcal{H}} \frac{\mathbf{P}\left[H = d_i\right]}{\sum_{d_j \in \mathcal{H}} d_j \mathbf{P}\left[H = d_j\right]} \\
&= \frac{1}{\mathbf{E}\left[H\right]}.
\end{aligned} \tag{31}$$

Define G to be the duration of an active connection, and note that $\mathbf{E}\left[\text{idle period}\right] = 1/P_{ar}$. Then, by a simple argument from renewal theory we get

$$\mathbf{P}\left[\text{active}\right] = \frac{\mathbf{E}\left[G\right]}{\mathbf{E}\left[G\right] + 1/P_{ar}}.$$

Given the RTT d and the initial workload x, then

$$\mathbf{E}\left[G | \text{RTT is } d \text{ and initial workload is } x\right] = \frac{x}{T(d, f(q^\star))}, \tag{32}$$

where $T(d, f(q^\star))$ is the average throughput of a TCP flow with RTT of d and packet marking rate of $f(q^\star)$. From (A5) we make use of the following approximation for the average throughput of a long-lived TCP flow [5], for some constant K,

$$T(d, f(q^\star)) \approx \frac{K}{d\sqrt{f(q^\star)}},$$

which implies that

$$\mathbf{E}\left[G|\text{delay is } d \text{ and initial workload is } x\right] \approx \frac{xd\sqrt{f(q^\star)}}{K}.$$

Since the initial workload and the delay are independent, we have

$$\mathbf{E}\left[G\right] = \frac{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[H\right]\sqrt{f(q^\star)}}{K}.$$

Therefore,

$$\begin{aligned}\mathbf{P}\left[\text{active}\right] &= \frac{\frac{\mathbf{E}[F_{ar}]\mathbf{E}[H]\sqrt{f(q^\star)}}{K}}{\frac{\mathbf{E}[F_{ar}]\mathbf{E}[H]\sqrt{f(q^\star)}}{K} + 1/P_{ar}} \\ &= \frac{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[H\right]\sqrt{f(q^\star)}}{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[H\right]\sqrt{f(q^\star)} + K/P_{ar}} \end{aligned} \tag{33}$$

Finally, in order to compute (28), we need to calculate $\mathbf{E}\left[\min(W^\star, X^\star)|\text{active}\right]$ which can be approximated by $\mathbf{E}\left[W^\star|\text{active}\right]$ under (A5), *i.e.*, , $\mathbf{E}\left[F\right] >> \mathbf{E}\left[W^\star\right]$.

$$\begin{aligned}&\mathbf{E}\left[W^\star|\text{active}\right] \\ &= \sum_{d_i \in \mathcal{H}} \mathbf{E}\left[W^\star|d^\star = d_i, \text{active}\right]\mathbf{P}\left[d^\star = d_i|\text{active}\right] \\ &= \sum_{d_i \in \mathcal{H}} \frac{K}{\sqrt{f(q^\star)}} \frac{d_i \mathbf{P}\left[H = d_i\right]}{\sum_{d_j \in \mathcal{H}} d_j \mathbf{P}\left[H = d_j\right]} \\ &= \frac{K}{\sqrt{f(q^\star)}} \end{aligned} \tag{34}$$

Combining (28), (31), (33) and (34), we get

$$\mathbf{E}\left[A^\star\right] \approx \frac{K\mathbf{E}\left[F_{ar}\right]}{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[H\right]^2\sqrt{f(q^\star)} + K\mathbf{E}\left[H\right]/P_{ar}}.$$

If $f(q^\star) \in (0, 1)$, it is necessary that

$$\begin{aligned}C &= \mathbf{E}\left[A^\star\right] \\ &\approx \frac{K\mathbf{E}\left[F_{ar}\right]}{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[H\right]\sqrt{f(q^\star)} + K/P_{ar}}. \end{aligned}$$

After some simple algebras, we can solve for $f(q^\star)$:

$$f(q^\star) = K^2\left(\frac{1}{C\mathbf{E}\left[H\right]} - \frac{1}{P_{ar}\mathbf{E}\left[F_{ar}\right]}\right)^2. \tag{35}$$

If f is invertible, then

$$q^\star = f^{-1}\left(K^2\left(\frac{1}{C\mathbf{E}\left[H\right]} - \frac{1}{P_{ar}\mathbf{E}\left[F_{ar}\right]}\right)^2\right). \tag{36}$$

A numerical example that validates our analysis is given in Section VI. This simple formulation can be used as a guideline on how to design the feedback probability function to control the queue size at the steady-state.

V. Discussion

Theorem 1 shows that the dynamics of the queue at time t, denoted by $Q^{(N)}(t)$, can be approximated by $Nq(t)$ with $q(t)$ determined via a simple deterministic recursion, which is independent of the number of sessions. The offered traffic into the network during the timeslot, $A^{(N)}(t)$, can also be approximated by $N \cdot \mathbf{E}\left[A(t)\right]$. These approximations become more accurate as the number of sessions becomes large, and the computational complexity does not depend on N. The limiting model is therefore "scalable" as it does not suffer from the explosion of state space, nor does it require any ad-hoc assumptions.

Theorem 1 also shows that the dependency between each session becomes negligible under a large number of sessions, *i.e.,* "RED breaks the global synchronization when the number of sessions is large."

Although the sequence $\{(q(t), \hat{q}(t), \mathbf{Y}(t)),\ t = 0, 1, \ldots\}$ is a time-homogeneous Markov chain, we do not address here the existence of the steady-state when $t \to \infty$ as complications arise due the fact that the first two components are degenerate (*i.e.,* deterministic). However, we note that the numerical calculations for the limiting model are simple. The number of steps required for the calculation for each time step is independent of N.

It is also interesting to see that the steady-state marking probability and the average queue size are affected by the round-trip delay only through the *mean* round-trip delay as suggested in (35) and (36), respectively. However, we conjecture that the variance of the round-trip delay will play a role in the magnitude of the queue fluctuations even though the mean queue size is determined only through the average delay. This is demonstrated in the simulation results in the next section.

VI. Numerical Example

This section presents a numerical example to study the behavior of the queue size per flow.

1) Example (i): The system and control parameters are set as follows:

$$C = 1 \text{ packet/timeslot},\ q^N_{\min} = 2 \cdot N,\ q^N_{\max} = 20 \cdot N, p_{\max} = 0.2$$

The initial values are set to $Q^{(N)}(0) = W_i^{(N)}(0) = X_i^{(N)}(0) = \beta_i^{(N)}(0) = d_i^{(N)}(0) = 0$ and $S_i^{(N)}(0) = M_i^{(N)}(0) = 1$ for $i \in \mathcal{N}$ at the beginning. The variables evolve according to the dynamics outlined in the previous section. The workload $F_i(t) \sim Geometric(p), i \in \mathcal{N}$ and $t = 0, 1, \ldots$ where $p = 0.001$, *i.e.,* , $\mathbf{E}\left[F_i(t)\right] = 1,000$ packets. The idle periods of sessions are geometrically distributed with a mean of 20 timeslots. The receiver advertised window size $W_{\max}$ is set to 64. The exponential average parameter α is set to 0.01.

First, we simulate the dynamics when the random round-trip delay $H_i(t)$ is uniformly distributed on the set $\mathcal{H} = \{2, \ldots, 6\}$. Figure 1 and 2 plot the evolution of the queue size per flow and the average queue size per flow, respectively, with the number of sessions $N = 100, 500, 1000, 5000,$ and 10000. As expected, the oscillation in the queue size per flow decreases with increasing N. Given the parameters used in this example with $K = \sqrt{3/2}$ [5], we have

$$q^\star \approx f^{-1}\left(K^2\left(\frac{1}{C\mathbf{E}\left[H\right]} - \frac{1}{P_{ar}\mathbf{E}\left[F_{ar}\right]}\right)^2\right)$$

$$
\begin{aligned}
&= f^{-1}\left(\frac{3}{2}\left(\frac{1}{(2+6)/2}-\frac{1}{0.05\cdot 1000}\right)^2\right)\\
&= 9.15.
\end{aligned}
$$

Hence, as can be seen from Figure 1, the steady-state queue size is close to the value predicted by the model.

We demonstrate the role of the variability in the round-trip in the next example. In this case, the round-trip delay of each connection is either 2 or 6 with equal probability, *i.e.*, ,

$$\mathbf{P}\left[H_i(t)=2\right]=\mathbf{P}\left[H_i(t)=6\right]=0.5,\quad t=0,1,\ldots, \tag{37}$$

which has the maximum variance of any distribution in $\mathcal{H}$ that still has the mean value of 4. The rest of the setup is identical to the previous case.

Figure 3 and 4 plot the evolution of the queue size per flow and the average queue size per flow, respectively. While the queue size converges to the same value as in the previous example, notice that (i) the magnitude of the fluctuation for the same number of user is greater when the round-trip distribution is Bernoulli, and (ii) the convergence to steady-state is slower (in time) for Bernoulli distribution. This clearly demonstrates that while the steady-state mean queue size is determined only by the distribution, the transient behavior and the magnitude of fluctuation is affected by the variability in the round-trip delays of the flows.

2) Example (ii): In the second example, we change the capacity per flow, workload and the idle period distribution to be as follows:

$$C=0.6 \text{ packet/timeslot}$$

The workload $F_i(t)\sim Geometric(p), i\in\mathcal{N}$ and $t=0,1,\ldots$ where $p=0.005$, i.e., $\mathbf{E}\left[F_i(t)\right]=200$ packets. The idle periods of sessions are geometrically distributed with a mean of 5 timeslots. The rest of the parameters are identical to Example (i).

Again, we simulate the dynamics when the random round-trip delay $H_i(t)$ is uniformly distributed on the set $\mathcal{H}=\{2,\ldots,10\}$ so the average round-trip delay equals 6 timeslots. Figure 5 and 6 plot the evolution of the queue size per flow and the average queue size per flow, respectively, with the number of sessions $N=100, 500, 1000, 5000,$ and 10000. Figure 7 and 8 plot the evolution when $H_i(t)$ is a Bernoulli rv with the following distribution:

$$\mathbf{P}\left[H_i(t)=2\right]=\mathbf{P}\left[H_i(t)=10\right]=0.5,\quad t=0,1,\ldots; i=1,2,\ldots,N, \tag{38}$$

so that $\mathbf{E}\left[H_i(t)\right]=6$.

Given the parameters used in this example with $K=\sqrt{3/2}$ [5], we have

$$
\begin{aligned}
q^\star &\approx f^{-1}\left(K^2\left(\frac{1}{C\mathbf{E}\left[H\right]}-\frac{1}{P_{ar}\mathbf{E}\left[F_{ar}\right]}\right)^2\right)\\
&= f^{-1}\left(\frac{3}{2}\left(\frac{1}{(0.6)(6)}-\frac{1}{0.2\cdot 200}\right)^2\right)\\
&= 10.63,
\end{aligned}
$$

which is close to the steady-state queue level in both simulations. Therefore, this verifies that the steady-state average queue size depends only on the mean round-trip delay.

From the simulation results, we draw the conclusion that (i) the oscillation in the queue size per flow decreases with increasing N, (ii) the level of the queue size at steady-state depends only on the mean RTT, and (iii) the magnitude of the queue fluctuation depends on the distribution of the RTT.

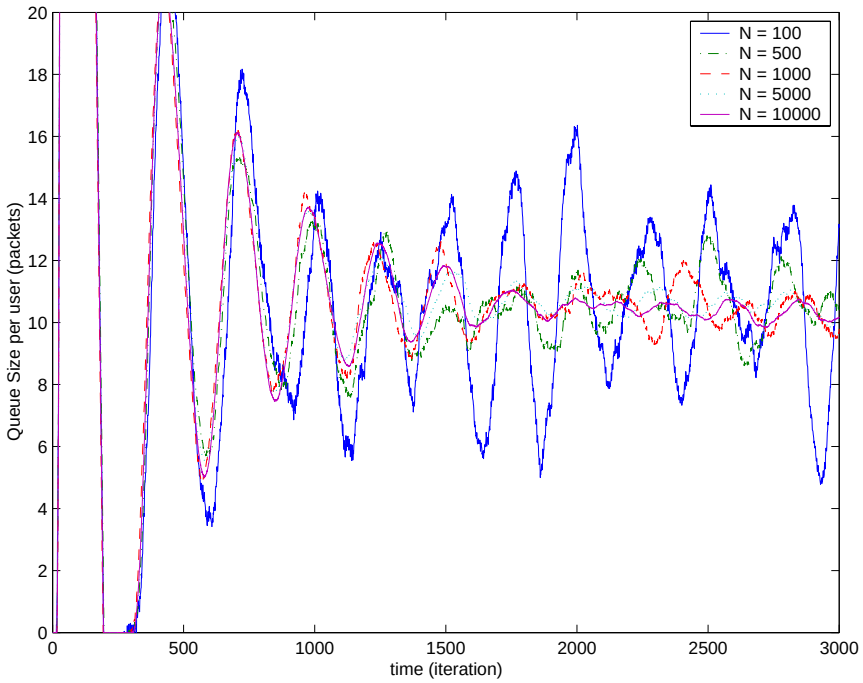

Fig. 1. Example (i) : Evolution of queue size per flow when the round-trip is uniform.

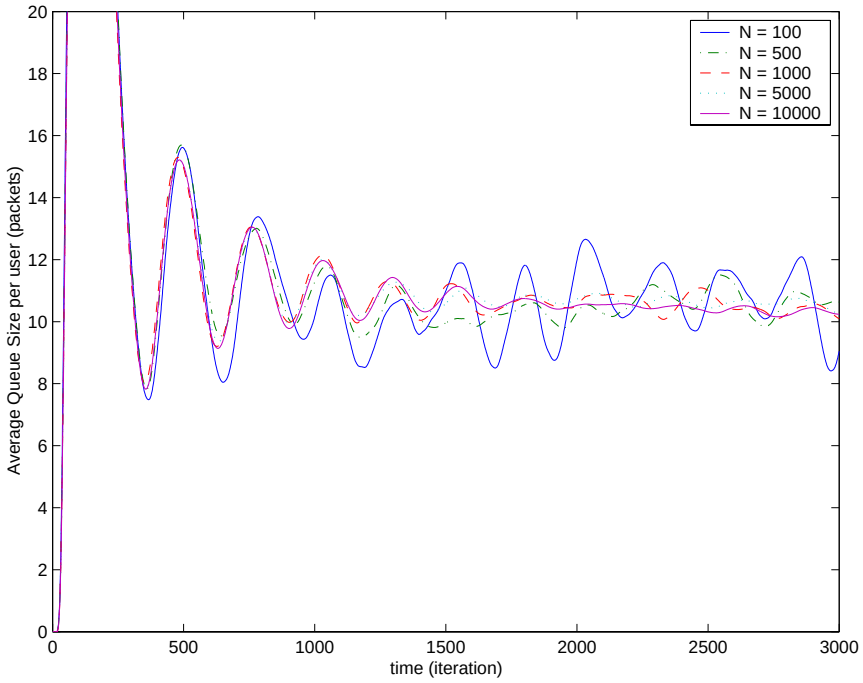

Fig. 2. Example (i) : Evolution of average queue size per flow when the round-trip is uniform.

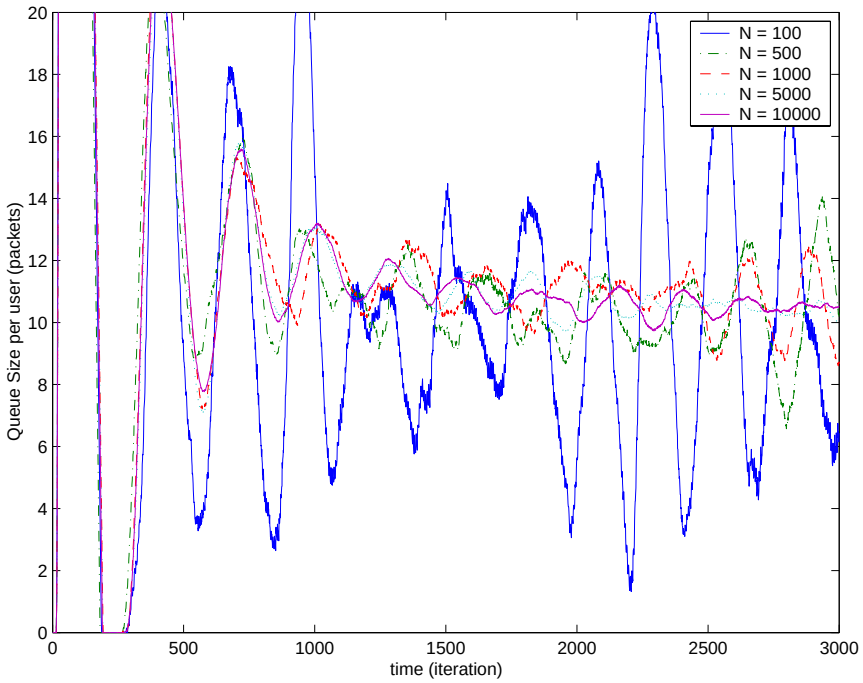

Fig. 3. Example (i) : Evolution of queue size per flow when the round-trip is Bernoulli.

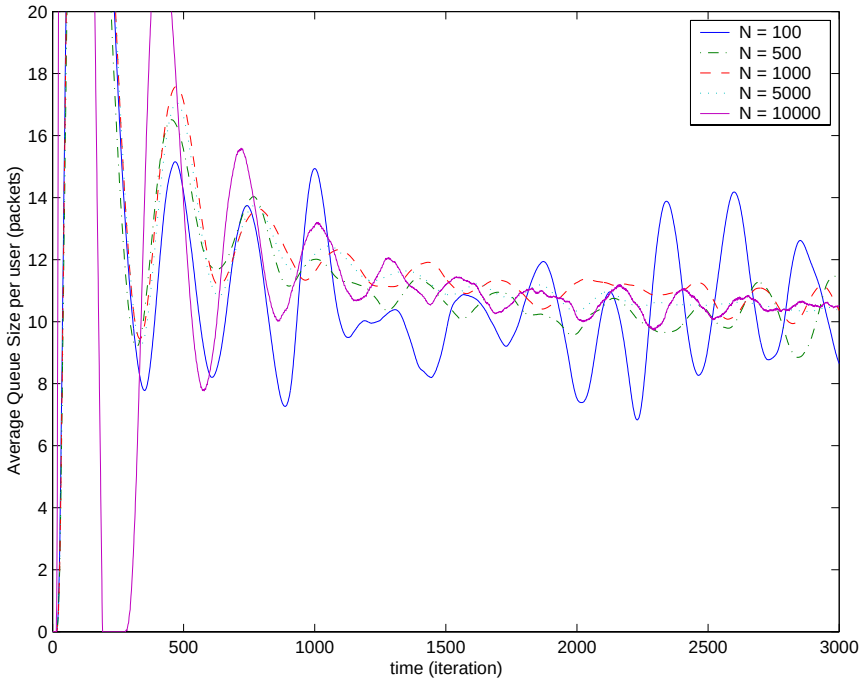

Fig. 4. Example (i) : Evolution of average queue size per flow is Bernoulli.

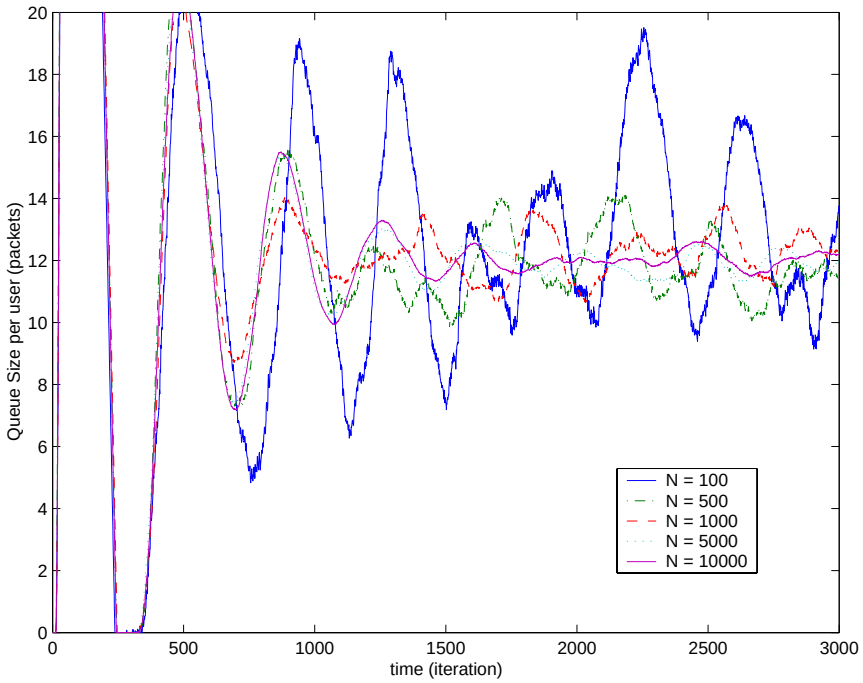

Fig. 5. Example (ii) : Evolution of queue size per flow when the round-trip is uniform.

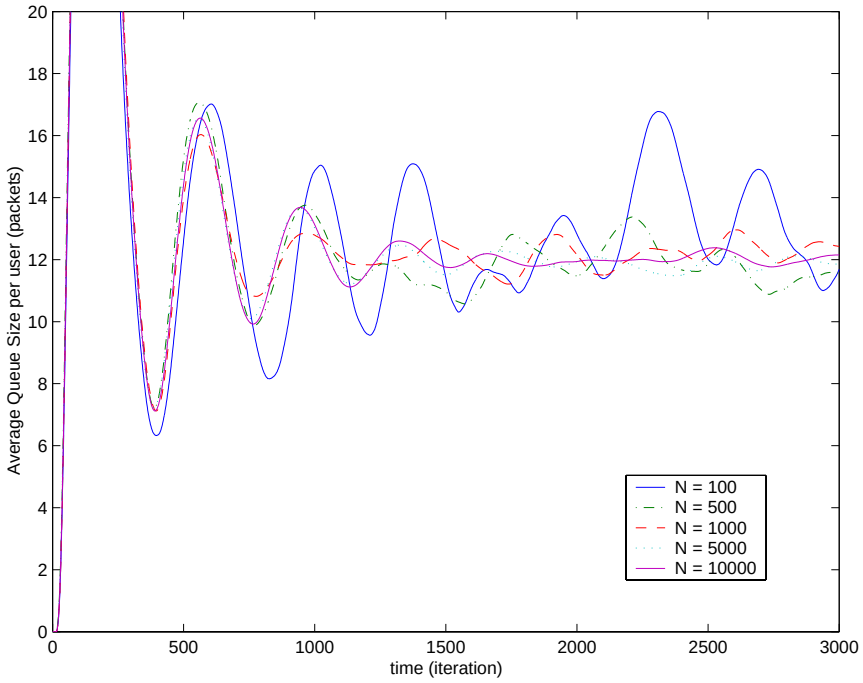

Fig. 6. Example (ii) : Evolution of average queue size per flow when the round-trip is uniform.

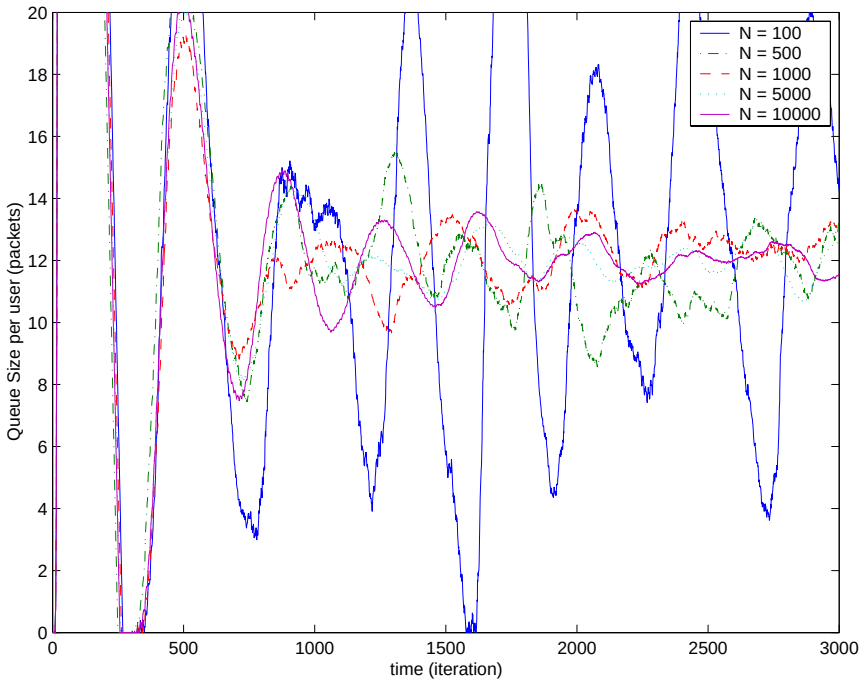

Fig. 7. Example (ii) : Evolution of queue size per flow when the round-trip is Bernoulli.

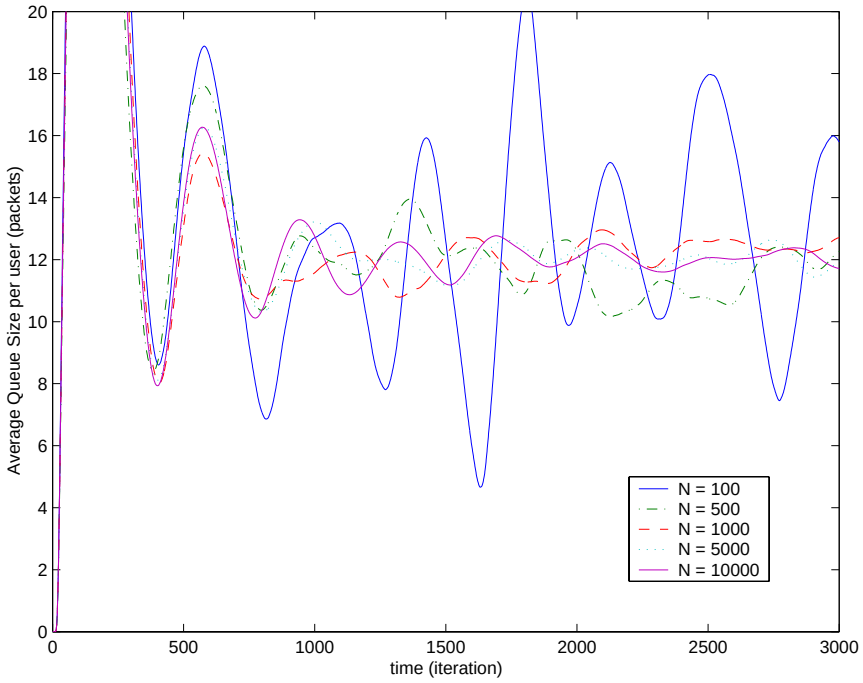

Fig. 8. Example (ii) : Evolution of average queue size per flow is Bernoulli.

VII. NS-2 Simulation results

In this section we verify our analysis by a more realistic event-driven *NS-2* simulations. In the simulation we gradually vary the number of sessions from 25 to 1,000, and study the queue dynamics. The system parameters used in the simulation are scaled with the number of sessions N as follows: the bottleneck link capacity $C^{(N)} = 0.24 \cdot N$ Mbps, the bottleneck buffer $B^{(N)} = 25 \cdot N$ packets. The bottleneck RED gateway with ECN option enabled is configured with the following marking probability function $f^{(N)}(x) = f(N^{-1}x)$ (*i.e.,* a scaling similar to assumption (A1)) where it corresponds to the following parameters setting in *NS-2*: $q_{min}^{(N)} = 2 \cdot N, q_{max}^{(N)} = 10 \cdot N, p_{max} = 0.1$ with the gentle mode enabled. The receiver advertised window W_{max} is set to 64 packets and the packet size is fixed to 1,000 bytes. The exponential averaging weight of the RED gateway is set to $0.02/N$ in order to have a similar time constant in all cases. A session generates a workload that is exponentially distributed with a mean of 100 packets, and the interarrival times of the new workloads for each connection are exponentially distributed with a mean of 3.3 seconds. When a session runs out of data to transfer, it terminates the TCP connection. A new TCP connection is initiated by the session when the next workload arrives for the session. We also enable the *drop_front_* option, *i.e.,* the RED gateway marks the packet at the front of the queue rather than the packet that has just arrived, in order to reduce the feedback delay of the marks to the TCP sender.

The simulation results are obtained with two types of session round-trip delay distributions. First, the round-trip propagation delays of the sessions are randomly selected uniformly from $[52, 121.5]$ ms, with a mean of 87 ms. The simulation result is shown in Figure 9. Next, the round-trip propagation delays of the sessions are randomly selected to be either 52 or 121.5 ms with equal probability (*i.e.,* i.i.d. Bernoulli rv). This delay distribution also has the mean of 87ms but with much higher variance. Figure 10 shows the simulation result from this setting.

Notice that the fluctuations in the normalized queue level decrease as the number of sessions N increases. Furthermore, observations on steady-state queue level and fluctuations are all in agreement with the conclusion from Section VI.

VIII. Conclusions

In this paper, we have developed a stochastic model which includes the session, and network dynamics of the interaction between RED and many TCP flows with heterogeneous RTTs. As the number of sessions become large, the model can be approximate by a simple stochastic recursion which is independent of the number of sessions, *i.e.,* the limiting model is scalable. A fixed-point analysis of the steady-state queue size is also demonstrated and is verified by Monte-Carlo simulations of the model. NS-2 simulation results also confirm the qualitative findings of the analysis.

References

[1] S. Floyd. TCP and explicit congestion notification. *ACM Computer Communication Review*, 24:10–23, October 1994.

[2] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1993.

[3] Van Jacobson. Congestion avoidance and control. In *Proceedings of SIGCOMM'88 Symposium*, pages 314–332, August 1988.

[4] Steven H. Low, Fernando Paganini, Jiantao Wang, Sachin Adlakha, and John C. Doyle. Dynamics of TCP/RED and a scalable control. In *Proceedings of IEEE INFOCOM*, 2002.

[5] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: a simple model and its empirical validation. *IEEE/ACM Transactions on Networking*, 8(2):133–145, April 2000.

[6] Vern Paxson and Sally Floyd. Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3:226–244, 1995.

[7] Peerapol Tinnakornsrisuphap, Richard J. La, and Armand M. Makowski. Characterization of general TCP traffic under a large number of flows regime. Technical report, Institute for Systems Research, University of Maryland, 2002.

[8] Peerapol Tinnakornsrisuphap and Armand M. Makowski. Limit behavior of ECN/RED gateways under a large number of TCP flows. In *Proceedings of IEEE INFOCOM*, San Francisco, CA, April 2003.

[9] L. Zhang, S. Shenker, and D. Clark. Observations on the dynamics of a congestion control algorithm: The effects of two-way traffic. In *Proceedings of ACM SIGCOMM*, pages 133–147, September 1991.

[10] Alan F. Karr, *Probability*, Springer-Verlag, 1993.

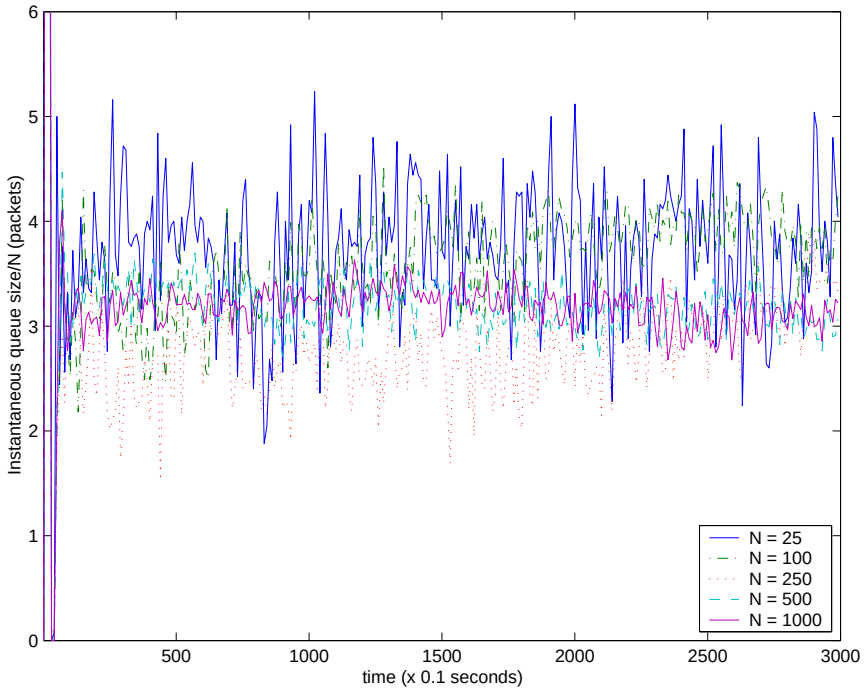

Fig. 9. Queue dynamics of the NS-2 simulation where the round-trip distribution is uniform

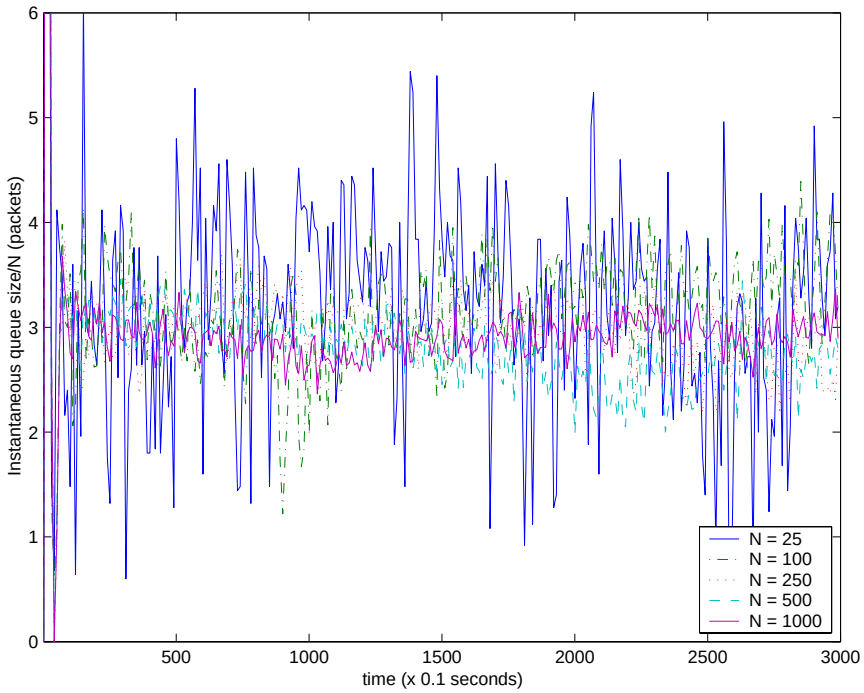

Fig. 10. Queue dynamics of the NS-2 simulation where the round-trip distribution is Bernoulli

APPENDIX I
PROOF OF THEOREM 1

A. Some simple and useful facts

Fix $i = 1, \ldots, N$ and consider an arbitrary *bounded* mapping $g : \not{\mathbb{Z}}_+^6 \to \mathbb{R}$: Through careful case analysis, it follows from that

$$
\begin{aligned}
& g(\mathbf{Y}_i^{(N)}(t+1)) \qquad (39)\\
& = \mathbf{1}\left[X_i^{(N)}(t) = 0\right] g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] H_i(t+1), 0, 1\right)\\
& + \mathbf{1}\left[X_i^{(N)}(t) > A_i^{(N)}(t) > 0\right] g(W_i^{(N)}(t), X_i^{(N)}(t) - A_i^{(N)}(t),\\
& \quad \prod_{j=1}^{A_i^{(N)}(t)} M_{i,j}(t+1) S_i^{(N)}(t), d_i^{(N)}(t), 1, \prod_{j=1}^{A_i^{(N)}(t)} M_{i,j}(t+1))\\
& + \mathbf{1}\left[0 < X_i^{(N)}(t) \le A_i^{(N)}(t)\right] g(0, 0, 1, d_i^{(N)}(t), 1, \prod_{j=1}^{A_i^{(N)}(t)} M_{i,j}^{(N)}(t+1))\\
& + \mathbf{1}\left[X_i^{(N)}(t) > 0, \beta_i^{(N)}(t) = d_i^{(N)}(t) - 1\right] g(W_{i,new}^{(N)}(t+1), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), d_i^{(N)}(t), M_i^{(N)}(t))\\
& + \mathbf{1}\left[X_i^{(N)}(t) > 0, \beta_i^{(N)}(t) < d_i^{(N)}(t) - 1\right] g(W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), \beta_i^{(N)}(t) + 1, M_i(t))
\end{aligned}
$$

where

$$
\begin{aligned}
& g(W_{i,new}^{(N)}(t+1), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), d_i^{(N)}(t), M_i^{(N)}(t))\\
= \;& M_i^{(N)}(t) S_i^{(N)}(t) F_g^1\left(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)\right)\\
+ \;& M_i^{(N)}(t)(1 - S_i^{(N)}(t)) F_g^2\left(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)\right)\\
+ \;& (1 - M_i^{(N)}(t)) F_g^3(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)). \qquad (40)
\end{aligned}
$$

The $\not{\mathbb{Z}}_+^3 \to \mathbb{R}$ mappings F_g^1, F_g^2 and F_g^3 are associated with g and defined as follows:

$$
\begin{aligned}
F_g^1\left(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)\right) &= g\left(\min\left(2W_i^{(N)}(t) \vee 1, W_{\max}\right), X_i^{(N)}(t), 1, d_i^{(N)}(t), d_i^{(N)}(t), 1\right)\\
F_g^2\left(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)\right) &= g\left(\min\left(W_i^{(N)}(t) + 1, W_{\max}\right), X_i^{(N)}(t), 0, d_i^{(N)}(t), d_i^{(N)}(t), 1\right)\\
F_g^3\left(W_i^{(N)}(t), X_i^{(N)}(t), d_i^{(N)}(t)\right) &= g\left(\lceil \frac{W_i^{(N)}(t)}{2} \rceil, X_i^{(N)}(t), 0, d_i^{(N)}(t), d_i^{(N)}(t), 0\right) \qquad (41)
\end{aligned}
$$

Let $\mathcal{F}_t$ denote the σ-field generated by the rvs $\{Q^{(N)}(0), W_i^{(N)}(0), X_i^{(N)}(0), U_i(s), F_i(s), H_i(s), V_i(s), V_{i,j}(s)$, $i, j = 1, 2, \ldots ; s = 1, \ldots, t\}$ with the rvs $Q^{(N)}(t)$ and $\mathbf{Y}_i^{(N)}(t)$ ($i = 1, \ldots, N$) being all $\mathcal{F}_t$-measurable, it holds under the enforced independence assumptions that

$$
\mathbf{E}\left[M_{i,j}^{(N)}(t+1)|\mathcal{F}_t\right] = 1 - f^{(N)}(\hat{Q}^{(N)}(t)), \quad j = 1, 2, \ldots
$$

so that

$$
\mathbf{E}\left[\prod_{i=1}^{A(t)} M_{i,j}^{(N)}(t+1)|\mathcal{F}_t\right] = Z_i^{(N)}(t) \qquad (42)
$$

by conditional independence, where we have set

$$Z_i^{(N)}(t) = \left(1 - f^{(N)}(\hat{Q}^{(N)}(t))\right)^{A_i^{(N)}(t)}. \tag{43}$$

It is now clear that

$$\prod_{i=1}^{A(t)} M_{i,j}^{(N)}(t+1) =_{st} \mathbf{1}\left[V_i(t+1) \leq Z_i^{(N)}(t)\right]. \tag{44}$$

It readily follows from (39) that

$$\begin{aligned}
&\mathbf{E}\left[g(\mathbf{Y}_i^{(N)}(t+1))|\mathcal{F}_t\right] \\
&= \mathbf{1}\left[X_i^{(N)}(t) = 0\right] \mathbf{E}\left[g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] H_i(t+1), 0, 1\right)\right] \\
&+ \mathbf{1}\left[X_i^{(N)}(t) > A_i^{(N)}(t) > 0\right] \\
&\quad \times [Z_i^{(N)}(t) g(W_i^{(N)}(t), X_i^{(N)}(t) - A_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), 1, 1) \\
&\quad + (1 - Z_i^{(N)}(t)) g(W_i^{(N)}(t), X_i^{(N)}(t) - A_i^{(N)}(t), 0, d_i^{(N)}(t), 1, 0)] \\
&+ \mathbf{1}\left[0 < X_i^{(N)}(t) \leq A_i^{(N)}(t)\right] \\
&\quad \times [Z_i^{(N)}(t) g(0, 0, 1, d_i^{(N)}(t), 1, 1) + (1 - Z_i^{(N)}(t)) g(0, 0, 1, d_i^{(N)}(t), 1, 0)] \\
&+ \mathbf{1}\left[X_i^{(N)}(t) > 0, \beta_i^{(N)}(t) = d_i^{(N)}(t) - 1\right] g(W_{i,new}^{(N)}(t+1), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), d_i^{(N)}(t), M_i^{(N)}(t)) \\
&+ \mathbf{1}\left[X_i^{(N)}(t) > 0, \beta_i^{(N)}(t) < d_i^{(N)}(t) - 1\right] g(W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), \beta_i^{(N)}(t) + 1, M_i(t)) \\
&= F_g(Z_i^{(N)}(t), W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), d_i^{(N)}(t), \beta_i^{(N)}(t), M_i^{(N)}(t))
\end{aligned} \tag{45}$$

where the mapping $F_g : [0,1] \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0,1\} \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0,1\} \to \mathbb{R}$ is associated with g through

$$\begin{aligned}
&F_g(z, w, x, s, d, b, m) \\
&= \mathbf{1}[x = 0] \mathbf{E}\left[g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] H_i(t+1), 0, 1\right)\right] \\
&+ \mathbf{1}[x > \min(w, x) \mathbf{1}[b \geq d > 0]] \\
&\quad \times [z g(w, x - \min(w, x) \mathbf{1}[b \geq d > 0], s, d, 1, 1) + (1 - z) g(w, x - \min(w, x) \mathbf{1}[b \geq d > 0], 0, d, 1, 0)] \\
&+ \mathbf{1}[0 < x \leq \min(w, x) \mathbf{1}[b \geq d > 0]] [z g(0, 0, 1, d, 1, 1) + (1 - z) g(0, 0, 1, d, 1, 0)] \\
&+ \mathbf{1}[x > 0, b = d - 1] g_{new}(w, x, s, d, m) \\
&+ \mathbf{1}[x > 0, b < d - 1] g(w, x, s, d, b + 1, m)
\end{aligned} \tag{46}$$

where

$$g_{new}(w, x, s, d) = ms F_g^1(w, x, d) + m(1 - s) F_g^2(w, x, d) + (1 - m) F_g^3(w, x, d) \tag{47}$$

We note that $\mathbf{E}\left[g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] H_i(t+1), 0, 1\right)\right]$ always exists and is finite because the mapping g is bounded. Furthermore, the mapping F_g is continuous with respect to the product topology on $[0,1] \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0,1\} \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0,1\} \to \mathbb{R}$.

Upon taking expectations on both sides of (45) we see that

$$\mathbf{E}\left[g\left(Y_i^{(N)}(t+1)\right)\right] = \mathbf{E}\left[F_g\left(Z_i^{(N)}(t), Y_i^{(N)}(t)\right)\right]. \tag{48}$$

B. A Weak Law of Large Numbers

We introduce the following terminology to facilitate the discussion: For each $t = 0, 1, \ldots$, the statements **[A:t]**, **[B:t]**, **[C:t]** and **[D:t]** below refer to the following convergence statements:

[A:t] For some non-random constant $q(t), \hat{q}(t)$, it holds that

$$\begin{aligned} \frac{Q^{(N)}(t)}{N} &\xrightarrow{P}_N q(t); \\ \text{and } \frac{\hat{Q}^{(N)}(t)}{N} &\xrightarrow{P}_N \hat{q}(t) \end{aligned} \tag{49}$$

[B:t] For some $\{0, 1, \ldots, W_{\max}\}$-valued rv $W(t)$, non-negative integer-valued rv $X(t), \beta(t), d(t)$, and $\{0, 1\}$-valued rv $S(t), M(t)$, it holds that

$$\mathbf{Y}_1^{(N)}(t) \Rightarrow_N \mathbf{Y}(t) = (W(t), X(t), S(t), d(t), \beta(t), M(t)); \tag{50}$$

[C:t] For any integer $I = 1, 2, \ldots$, the rvs $\{\mathbf{Y}_i^{(N)}(t),\ i = 1, \ldots, I\}$ become asymptotically independent with large N as described by (20) where $\mathbf{Y}(t)$ are the rvs occurring in **[B:t]**;

[D:t] For any bounded mapping $g : \mathbb{Z}_+^6 \to \mathbb{R}$, the convergence (17) holds with $\mathbf{Y}(t)$ the rvs occurring in **[B:t]**. Moreover, if the file arrival distribution has a finite second moment, then the convergence (19) also holds.

With the help of a series of lemmas, we shall prove the validity of the statements **[A:t]**–**[D:t]** for all $t = 0, 1, \ldots$. We do so by induction on t and in the process we establish Theorem 1.

Lemma 2: Under (A1), if **[A:t]** *and* **[B:t]** *hold for some* $t = 0, 1, \ldots$, *then* **[B:t+1]** *holds with* $\mathbf{Y}(t+1)$ *related in distribution to* $\mathbf{Y}(t)$ *by (24).*

Proof: Together the convergence **[A:t]** and **[B:t]** imply [10, Thm. 5.28, p. 150] the joint convergence

$$(N^{-1}\hat{Q}^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \Rightarrow_N (\hat{q}(t), \mathbf{Y}(t)). \tag{51}$$

Next the continuity of the mapping f implies that of $(y, w, x, b, d) \to (1 - f(y))^{\min(w,x)\mathbf{1}[b \geq d]}$ on $\mathbb{R}_+ \times [0, \infty) \times [0, \infty) \times \mathbb{Z}_+ \times \mathbb{Z}_+$, so that

$$(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \Rightarrow_N (Z(t), \mathbf{Y}(t)) \tag{52}$$

by the Continuous Mapping Theorem [10, Thm. 5.29, p. 150] with

$$Z(t) = (1 - f(\hat{q}(t)))^{\min(W(t), X(t))\mathbf{1}[\beta(t) \geq d(t)]}.$$

Consider (48) for any *bounded* arbitrary mapping $g : \mathbb{Z}_+^6 \to \mathbb{R}$, and recall that the mapping F_g defined by (46) is continuous on $[0, 1] \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0, 1\} \times \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{0, 1\}$. Consequently, the Continuous Mapping Theorem can again be invoked to yield

$$F_g(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \Rightarrow_N F_g(Z(t), \mathbf{Y}(t)), \tag{53}$$

whence

$$\lim_{N \to \infty} \mathbf{E}\left[F_g(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t))\right] = \mathbf{E}\left[F_g(Z(t), \mathbf{Y}(t))\right] \tag{54}$$

by the Bounded Convergence Theorem [10, Thm. 4.16, p. 108]. Combining (48) and (54) we get

$$\lim_{N \to \infty} \mathbf{E}\left[g(\mathbf{Y}_1^{(N)}(t))\right] = \mathbf{E}\left[F_g(Z(t), \mathbf{Y}(t))\right] \tag{55}$$

and since the bounded mapping g is arbitrary, it follows immediately that

$$\mathbf{Y}_1^{(N)}(t+1) \quad \Rightarrow_N \quad \mathbf{Y}(t+1) = (W(t+1), X(t+1), S(t+1), d(t+1), \beta(t+1), M(t+1))$$

for some $\{1, \ldots, W_{\max}\}$–valued rv $W(t+1)$, non-negative integer-valued $X(t+1), \beta(t+1), d(t+1)$ and $\{0,1\}$-valued rv $S(t+1)$, and $M(t+1)$ with

$$\mathbf{E}\left[g\left(\mathbf{Y}(t+1)\right)\right] \quad = \quad \mathbf{E}\left[F_g\left(Z(t), Y(t)\right)\right]. \tag{56}$$

A moment of reflection and a comparison to the analysis in (45)-(48) will convince the reader that (56) is equivalent to (24). ∎

Lemma 3: Under (A1), if **[A:t]** *and* **[D:t]** *hold for some* $t = 0, 1, \ldots$*, then* **[A:t+1]** *also holds.*

Proof: From **[A:t]** and **[D:t]** (specifically, (18)), we conclude that

$$\frac{Q^{(N)}(t)}{N} - C + \frac{1}{N}\sum_{i=1}^{N} A_i^{(N)}(t) \quad \stackrel{P}{\rightarrow}_N \quad q(t) - C + \mathbf{E}\left[A(t)\right] \tag{57}$$

and the desired result is a simple consequence of the continuity of the function $x \rightarrow x^+$ as we note that since

$$\frac{Q^{(N)}(t+1)}{N} = \left[\frac{Q^{(N)}(t)}{N} - C + \frac{1}{N}\sum_{i=1}^{N} A_i^{(N)}(t)\right]^+$$

for all $N = 1, 2, \ldots$. Since $\frac{\hat{Q}^{(N)}(t)}{N} \stackrel{P}{\rightarrow}_N \hat{q}(t)$ from **[A:t]**, it is simple to see that

$$\begin{aligned} \frac{\hat{Q}^{(N)}(t+1)}{N} \quad &= \quad (1-\alpha)\frac{\hat{Q}^{(N)}(t)}{N} + \alpha\frac{Q^{(N)}(t+1)}{N} \\ &\stackrel{P}{\rightarrow}_N \quad (1-\alpha)\hat{q}(t) + \alpha q(t+1) \end{aligned}$$

The proof of Lemma 3 also shows that

$$\frac{Q^{(N)}(t+1)}{N} \stackrel{P}{\rightarrow}_N q(t+1)$$

and

$$\frac{\hat{Q}^{(N)}(t+1)}{N} \stackrel{P}{\rightarrow}_N \hat{q}(t+1)$$

with non-random $q(t+1)$ determined by (21) and $\hat{q}(t+1)$ determined by (22). ∎

Lemma 4: Under (A1)–(A2), if **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some* $t = 0, 1, \ldots$*, then* **[C:t+1]** *also holds.*

Proof: We first observe that for a fixed N, the vector $(\mathbf{Y}_i^{(N)}(t), i = 1, \ldots, N)$ are coupled only through the marking probability which depends only on $\hat{Q}^{(N)}(t)$. Fix a positive integer I. The rvs $V_1(t+1), \ldots, V_I(t+1)$ are i.i.d. $[0,1]$-uniform rvs which are independent of $\mathcal{F}_t$. Thus, we see that the rvs $Y_1^{(N)}(t+1), \ldots, Y_I^{(N)}(t+1)$ are mutually independent given $\mathcal{F}_t$. Consequently, for arbitrary bounded mappings $g_1, \ldots, g_I : \mathbb{Z}_+^6 \rightarrow \mathbb{R}$, we get

$$\begin{aligned} \mathbf{E}\left[\prod_{i=1}^{I} g_i(\mathbf{Y}_i^{(N)}(t+1))|\mathcal{F}_t\right] \quad &= \quad \prod_{i=1}^{I} \mathbf{E}\left[g_i(\mathbf{Y}_i^{(N)}(t+1))|\mathcal{F}_t\right] \\ &= \quad \prod_{i=1}^{I} F_{g_i}(Z_i^{(N)}(t), \mathbf{Y}_i^{(N)}(t)) \end{aligned}$$

with the help of (45) and (46).

Now it follows from (20) in **[C:t]** that the joint convergence

$$\left(\mathbf{Y}_1^{(N)}(t),\ldots,\mathbf{Y}_I^{(N)}(t)\right) \quad \Rightarrow_N \quad (\mathbf{Y}_1(t),\ldots,\mathbf{Y}_I(t))$$

holds with limiting rvs $\mathbf{Y}_1(t)\ldots,\mathbf{Y}_I(t)$ which are i.i.d. random vectors each distributed according to $\mathbf{Y}(t)$. As in the proof of Lemma 2, the arguments leading to the convergence (53) also lead to

$$\begin{aligned}&(F_{g_1}(Z_1^{(N)}(t),\mathbf{Y}_1^{(N)}(t)),\ldots,F_{g_I}(Z_I^{(N)}(t),\mathbf{Y}_I^{(N)}(t)))\\ &\Rightarrow_N (F_{g_1}(Z_1(t),\mathbf{Y}_1(t)),\ldots,F_{g_I}(Z_I(t),\mathbf{Y}_I(t))\end{aligned}$$

where the limiting rvs $(Z_1(t),\mathbf{Y}_1(t)),\ldots,(Z_I(t),\mathbf{Y}_I(t))$ are i.i.d. rvs each distributed according to the pair $(Z(t),\mathbf{Y}(t))$. Therefore, by the Bounded Convergence Theorem,

$$\begin{aligned}\lim_{N\to\infty}\mathbf{E}\left[\prod_{i=1}^{I}g_i(\mathbf{Y}_i^{(N)}(t+1))\right] &= \lim_{N\to\infty}\mathbf{E}\left[\prod_{i=1}^{I}F_{g_i}(Z_i^{(N)}(t),\mathbf{Y}_i^{(N)}(t))\right]\\ &= \mathbf{E}\left[\prod_{i=1}^{I}F_{g_i}(Z_i(t),\mathbf{Y}_i(t))\right]\\ &= \prod_{i=1}^{I}\mathbf{E}\left[F_{g_i}(Z_i(t),\mathbf{Y}_i(t))\right]\\ &= \prod_{i=1}^{I}\mathbf{E}\left[g_i(\mathbf{Y}_i(t))\right] \qquad (58)\end{aligned}$$

where the last equality made use of the relation (56). The desired result **[C:t+1]** now follows from (58) given that the mappings $g_1,\ldots,g_I$ are arbitrary. ∎

Lemma 5: Under (A1)–(A2), if **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some* $t=0,1,\ldots$, *then* **[D:t]** *holds.*

Proof: Pick a mapping $g:\mathbb{Z}_+^6\to\mathbb{R}$. We begin by observing that under (A2) the rvs $\mathbf{Y}_i^{(N)}(t);\ i=1,\ldots,N$ are exchangeable. As a result, we get

$$\begin{aligned}&\mathrm{var}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N}g(\mathbf{Y}_i^{(N)}(t))\right]\\ =\ & N^{-2}\sum_{i=1}^{N}\mathrm{var}[g(\mathbf{Y}_i^{(N)}(t))]\\ +\ & N^{-2}\sum_{i,j=1,i\neq j}^{N}\mathrm{cov}[g(\mathbf{Y}_i^{(N)}(t)),g(\mathbf{Y}_j^{(N)}(t))]\\ =\ & N^{-1}\mathrm{var}[g(\mathbf{Y}_1^{(N)}(t))]+\frac{N-1}{N}\mathrm{cov}[g(\mathbf{Y}_1^{(N)}(t)),g(\mathbf{Y}_2^{(N)}(t))]. \qquad (59)\end{aligned}$$

Now let N go to infinity in (59): The validity of **[C:t]** and the Bounded Convergence Theorem already imply

$$\lim_{N\to\infty}\mathrm{cov}[g(\mathbf{Y}_1^{(N)}(t)),g(\mathbf{Y}_2^{(N)}(t))] \quad = \quad \mathrm{cov}[g(\mathbf{Y}_1(t)),g(\mathbf{Y}_2(t))]=0 \qquad (60)$$

by asymptotic independence. On the other hand,

$$\limsup_{N\to\infty}\mathrm{var}[g(\mathbf{Y}_1^{(N)}(t))]<\infty$$

since g is bounded.

Combining these observations we readily see that

$$\lim_{N\to\infty} \text{var}\left[\tfrac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t))\right] = 0,$$

whence, by Chebyshev's inequality,

$$\tfrac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t)) - \mathbf{E}\left[\tfrac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t))\right] \quad \stackrel{P}{\to}_N \quad 0. \tag{61}$$

This last convergence is equivalent to

$$\tfrac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t)) - \mathbf{E}\left[g(\mathbf{Y}_i^{(N)}(t))\right] \stackrel{P}{\to}_N 0$$

by exchangeability, and the desired convergence result (17) is now immediate once we remark under **[B:t]** that $\lim_{N\to\infty} \mathbf{E}\left[g(\mathbf{Y}_i^{(N)}(t))\right] = \mathbf{E}\left[g(\mathbf{Y}(t))\right]$. It is then straightforward to get (18) with $g(\mathbf{Y}_i^{(N)}(t)) = A_i^{(N)}(t) = \min(W_i^{(N)}(t), X_i^{(N)}(t))\mathbf{1}\left[\beta_i^{(N)}(t) \geq d_i^{(N)}(t)\right]$ and notice that $W_i^{(N)}(t)$ is bounded by $W_{\max}$.

Finally, if the file arrival distribution F has a finite second moment, the convergence (19) follows from the dominated convergence theorem on (60) with $g(\mathbf{Y}_i^{(N)}(t)) = X_i^{(N)}(t)$ and note that $X_i^{(N)}(t)$ is dominated by a random variable with the distribution F which has a finite second moment. ∎

We now conclude with a proof of Theorem 1: We first note that under (A1)-(A2) the statements **[A:t]**–**[D:t]** trivially hold for $t = 0$. Moreover, if **[A:t]**–**[C:t]** hold for some $t = 0, 1, \ldots$, then so do the statements **[D:t] [B:t+1]**, **[A:t+1]** and **[C:t+1]** by Lemma 5, Lemma 2, Lemma 3 and Lemma 4, respectively. Consequently, the statements **[A:t]**–**[D:t]** do hold for all $t = 0, 1, \ldots$ by induction and the validity of Claims (i)-(iii) of Theorem 1 is now established. The dynamics (21) is a byproduct of the proof of Lemma 3, while (24) is already contained in Lemma 2.