PH.D. THESIS

Dynamics of Random Early Detection Gateway Under A Large Number of TCP Flows

by Peerapol Tinnakornsrisuphap
Advisor:

CSHCN PhD 2004-1
(ISR PhD 2004-2)

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/

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2004		**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Dynamics of Random Early Detection Gateway Under A Large Number of TCP Flows	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space & Naval Warfare Systems Center,53560 Hull Street,San Diego,CA,92152-5001**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**161**	

ABSTRACT

Title of Dissertation: DYNAMICS OF RANDOM EARLY DETECTION GATEWAY UNDER A LARGE NUMBER OF TCP FLOWS

Peerapol Tinnakornsrisuphap, Doctor of Philosophy, 2004

Dissertation directed by: Professor Armand M. Makowski
Department of Electrical and Computer Engineering

While active queue management (AQM) mechanisms such as Random Early Detection (RED) are widely deployed in the Internet, they are rarely utilized or otherwise poorly configured. The problem stems from a lack of a tractable analytical framework which captures the interaction between the TCP congestion-control and AQM mechanisms. Traditional TCP traffic modeling has focused on "micro-scale" modeling of TCP, *i.e.,* detailed modeling of a single TCP flow. While micro-scale models of TCP are suitable for understanding the precise behavior of an individual flow, they are not well suited to the situation where a large number of TCP flows interact with each other as is the case in realistic networks.

In this dissertation, an innovative approach to TCP traffic modeling is proposed by considering the regime where the number of TCP flows competing for the bandwidth in the bottleneck RED gateway is large. In the limit, the queue size and the aggregate TCP traffic can be approximated by simple recursions which are independent of the number of flows. The limiting model is therefore scalable as it does not suffer from the state space explosion. The steady-state queue length and window distribution can be evaluated from well-known TCP models.

We also extend the analysis to a more realistic model which incorporates session-level dynamics and heterogeneous round-trip delays. Typically, ad-hoc assumptions are required to make the analysis for models with session-level dynamics tractable under a certain regime. In contrast, our limiting model derived here is compatible with other previously proposed models in their respective regime without having to rely on ad-hoc assumptions. The contributions from these additional layers of dynamics to the asymptotic queue are now crisply revealed through the limit theorems. Under mild assumptions, we show that the steady-state queue size depends on the file size and round-trip delay only through their mean values.

We obtain more accurate description of the queue dynamics by means of a Central Limit analysis which identifies an interesting relationship between the queue fluctuations and the random packet marking mechanism in AQM. The analysis also reveals the dependency of the magnitude of the queue fluctuations on the variability of the file size and round-trip delay. Simulation results supporting conclusions drawn from the limit theorems are also presented.

DYNAMICS OF RANDOM EARLY DETECTION GATEWAY UNDER A LARGE NUMBER OF TCP FLOWS

by

Peerapol Tinnakornsrisuphap

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor Armand M. Makowski, Chair
Professor John S. Baras
Professor Richard J. La
Professor Prakash Narayan
Professor A. Udaya Shankar

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Professor Armand M. Makowski, for his continuous support, patience and advice throughout the duration of my Ph.D. program. With his guidance, I have learned to strive for excellence.

My thanks also go to my co-advisor, Professor Richard J. La, for his time, efforts and insights on this research. In addition, I would like to thank the members of my dissertation committee: Professors John S. Baras, Prakash Narayan, and A. Udaya Shankar, for their comments and suggestions. Helpful comments and insights from Rajeev Agrawal, Steven H. Low, and the late Professor T. Y. Lee, are all appreciated.

I am thankful for funding support I have received from the following agencies: the Space and Naval Warfare Systems Center – San Diego under Contract No. N66001-00-C-8063 and Collaborative Technology Alliances (CTA) under Contract No. DAAD19-01-2-0011. My appreciation also extends to Research and Teaching Assistantship appointments from the Institute for Systems Research and Department of Electrical and Computer Engineering, University of Maryland.

On the personal side, a special thought goes to Rachanee Ungrangsi, whose encouraging messages during the writing of this dissertation are appreciated. Most importantly, I would like to express my gratitude to my family and friends. Without them, success means nothing.

TABLE OF CONTENTS

LIST OF FIGURES v

1 Introduction 1
1.1 Motivation and Goals 1
1.2 The Approach 5
1.3 Contributions 9
1.4 Thesis Organization 13
1.5 Notation 14

2 Literature Survey 15
2.1 Overview of Congestion-Control and AQM Mechanisms 15
2.2 Micro-scale Modeling of Congestion-Controlled Flows 23
2.3 Macro-scale Modeling of Congestion-Controlled Flows 30

3 A Rate-based Model for TCP/RED 35
3.1 The Model 36
3.2 The Weak Law of Large Numbers 38
3.3 Simulations of the Rate-based Model 41

4 A Window-based Model of Persistent TCP Population with Homogeneous Round-trip Delays 44
4.1 Model Description 46
4.2 The Asymptotics 50
4.3 Steady-state Regime 52
4.3.1 TCP throughput with fixed loss probability 54
4.3.2 Steady-state regime for the model in Section 4.1 56
4.3.3 Discussion 58
4.4 A Central Limit Theorem 60
4.5 Simulations 62
4.6 Simple Network Dimensioning 67

5 A Window-based Model with Session-level Dynamics and Heterogeneous Round-trip Delays 71
5.1 The Model . . . 73
5.2 The Asymptotics . . . 82
5.3 Limiting Cases . . . 85
5.4 Steady-State Regime . . . 86
5.5 A Central Limit Theorem . . . 91
5.6 Simulations of the Model with Session-level Dynamics and Variable Round-trip Delays . . . 95
5.7 A Proof of Lemma 1 . . . 102

6 Weak Laws of Large Numbers 105
6.1 A Weak Law of Large Numbers of a Generic Model . . . 105
6.2 Proof of Theorem 2 . . . 109
6.3 Proof of Theorem 4 . . . 111
6.4 A Proof of Theorem 6 . . . 115
6.5 A Note on the Distributional Recursions . . . 120

7 A Proof of Central Limit Theorem 121
7.1 A Proof of Proposition 2 . . . 123
7.2 A Key Decomposition . . . 125
7.3 The First Piece of the Puzzle . . . 127
7.4 The Delta Method . . . 128
7.5 A Proof of Proposition 3 . . . 129
7.6 A Proof of Proposition 1 . . . 131
7.7 Strengthening Claim (ii) of Theorem 2 . . . 132
7.8 A Proof of Proposition 4 . . . 134
7.9 A Proof of Proposition 5 . . . 136

8 Conclusions and Future Directions 139
8.1 Conclusions . . . 139
8.2 Future Directions . . . 142

Bibliography 146

LIST OF FIGURES

3.1 The normalized queue length of Simulation 1. 42
3.2 The average transmission rate per user of Simulation 1. 43

4.1 The mapping $\gamma \to \mathbf{E}[W^\gamma]$ when $W_{\max} = 5$ vs. the approximation (4.17). 58
4.2 The topology setup in NS simulation. 64
4.3 The normalized queue length of the ECN/RED gateway in NS simulation. 64
4.4 The normalized queue length of the model. 66
4.5 Sample standard deviation of the normalized queue. 67
4.6 The normalized queue length of the ECN/RED gateway in NS simulation with the marking probability function (4.32). 68
4.7 The normalized queue length of the model with the marking probability function (4.32). 69

5.1 Evolution of queue size per flow when the round-trip is uniform. . . 97
5.2 Evolution of queue size per flow when the round-trip is Bernoulli. . 98
5.3 Example (ii) : Evolution of queue size per flow when the round-trip is uniform. 99
5.4 Example (ii) : Evolution of queue size per flow when the round-trip is Bernoulli. 100
5.5 Queue dynamics of the NS-2 simulation where the round-trip distribution is uniform 102
5.6 Queue dynamics of the NS-2 simulation where the round-trip distribution is Bernoulli 103

Chapter 1

Introduction

1.1 Motivation and Goals

One of the key mechanisms for the operation of the best-effort service Internet is the congestion-control mechanism in the transmission control protocol (TCP) [26]. While there are several variations to the basic TCP congestion-control mechanism, they all have in common the *additive increase/multiplicative decrease* (AIMD) algorithm. This AIMD algorithm enables TCP congestion-control to be robust under diverse conditions. Unfortunately, the self-clocking feedback mechanism of the AIMD algorithm does introduce some additional complexities into the behavior of network traffic. There has been a number of studies trying to gain insights into this complex behavior. At the present, the relationship between the throughput of a single TCP flow and its round-trip and loss probability is fairly well understood [2, 40, 43, 48].

There are, however, certain aspects of TCP that are not well understood and which cannot be analyzed with the techniques in the aforementioned studies. This includes issues such as buffer behavior at a bottleneck router and the aggregate throughput when many TCP flows compete for the bandwidth of a

link. While one could in principle extend the aforementioned single flow models to attempt to answering these questions, the resulting models (which will be referred to later as *micro-scale models*) are not scalable. More specifically, since each flow is modeled in great details, the size of the state space for such models explodes when the number of flows becomes large. Consequently, not only is the analysis becoming intractable, even numerical calculations or simulations of such models are very complicated, computationally prohibitive, and would not provide additional advantages over full-scale simulation with existing simulation packages (*e.g.,* NS [45]). While one could make certain assumptions to simplify the analysis, it is not clear what are the irrelevant details that can be omitted while still providing a reasonably accurate analysis.

To make matters worse, recent developments in Active Queue Management (AQM) techniques have introduced additional complexity in transport protocols. The development of AQM originated from a well-known fact that with simple Tail-Drop gateways, the AIMD mechanism in TCP congestion-control leads to undesirable behavior, *i.e.,* global synchronization. When several TCP flows compete for bandwidth in a Tail-Drop gateway, it has been observed experimentally that packets from many flows are usually discarded simultaneously [63], resulting in a poor utilization of the network. AQM algorithms such as Random Early Detection (RED) [18] and Explicit Congestion Notification (ECN) [15] have been proposed to help alleviate this problem by randomly dropping/marking packets depending on queue size, thereby avoiding heavy congestion and preventing global synchronization. As can easily be imagined, the introduction of AQM further exacerbates the difficulty of understanding issues associated with buffer behavior and aggregate TCP traffic.

There have been attempts to model the interactions between TCP and AQM, but so far the analytically tractable models are either too crude or too simplistic (as to be discussed in the literature survey in Chapter 2).

Furthermore, Internet TCP traffic is composed of connections with diverse session dynamics and round-trip delays. The session dynamic of a flow depends on the type of applications. For example, applications such as FTP and Telnet, are relatively long-lived, while other types of applications are typically short-lived, *e.g.,* web browsing. Additional modeling difficulties also arise from the fact that the information (*i.e.,* marks on packets) from AQM to TCP flows is exchanged at different feedback rates depending on the round-trip delay of the flows. All of these features create major obstacles in deriving a scalable model which can cope with a large number of flows and still yields analytical insights into how to control such traffic. As a result, little work has been done on the problem of modeling the interaction between AQM and a *large number* of TCP flows with *session dynamics* and *variable round-trip delays*.

The problem of modeling network traffic and of understanding its impact on various performance metrics is not unique to TCP networks. Early traffic modeling efforts in telephony revealed the suitability of Poisson processes to model the time patterns in the stream of call requests to a telephone exchange. This culminated with the pioneering work of A.K. Erlang who proposed the $M|M|c|c$ model for dimensioning call systems [11]. Similarly, it is well-known that the multiplexed packet traffic generated by a large number of bursty data sources is well described by a Poisson process. It turns out in retrospect that this so-called "Poisson modeling" is easily justified by a celebrated limit theorem due to Palm [50] and Khintchine [33]. According to these results, under very weak

technical conditions, the superposition of point processes is a point process which converges to a Poisson (point) process when the number of superposed point processes grows large.

These success stories point to the possibility of systematically applying limit theorems in order to derive traffic models (which will be referred later as *macro-scale* models). The advantages of doing so are three-fold. First, model simplification typically occurs when applying limit theorems, thereby filtering out irrelevant details without relying on ad-hoc assumptions. Second, since limit theorems are central to the Theory of Probability, it is reasonable to expect the existence of suitable limit theorems (with very weak assumptions) which can be applied to the traffic model of interest. Finally, the resource allocation problem is interesting only in networks operating at high utilization such as when the number of users is large. In such a scenario, the limit behavior becomes more accurate as the number of users increases.

We would like to again stress that the development of a tractable analytical framework for the interaction between RED (and more generally AQM) and TCP is important in practice as well. While AQM mechanisms such as ECN and RED are widely implemented and supported in the new generation of routers, they are either rarely utilized or otherwise poorly configured due to the lack of an understanding of how to configure their parameters. In fact, May et al. [41] advise against the deployment of RED in spite of its benefits because a misconfigured RED can have detrimental effects on the network traffic. It is the objective of this thesis to show that simple and robust models can be derived in a systematic manner through limit theorems, and that such models can help improve the understanding of the dynamics and encourage the deployment of

well-engineered AQM mechanisms.

1.2 The Approach

As part of this program, we consider several discrete-time models of a bottleneck RED gateway with a large number of TCP flows competing for bandwidth resources. In this thesis, we mainly consider RED for the AQM mechanism because it is the simplest and most widely-deployed AQM mechanism. However, many of the conclusions drawn from the analytical results are applicable to models combining generic window-based congestion-control mechanisms and AQM schemes. Furthermore, the approach can be generalized to apply to a large class of models.

All the models considered here take into account detailed packet-level operations unlike other existing models in the literature (to be surveyed in Section 2.3), where the packet-level operations of TCP are ignored and only the evolution of the transmission rate is considered. It is important to make this distinction because mechanisms in TCP and AQM rely on packet-level operations. The omission of this level of details simplifies the analysis but at the expense of additional distortion in the results.

The models under investigation in this dissertation are organized around the behavior of the RED buffer contents observed at discrete epochs (which are indexed by $t = 0, 1, \ldots$). They differ through the specific mechanism used for generating incoming packets in response to the congestion that develops in the RED buffer shared by the TCP flows. Specifically, in the initial model (to be referred to as the *rate-based model* in Chapter 3) the packet transmission rate of

a source increases/decreases in response to random packet dropping at the RED gateway. This simple *ersatz* model is easy to understand and provides a complete conceptual picture of the system and the type of the results to be expected.

Later, we embellish this model by incorporating a more accurate additive increase/multiplicative decrease (AIMD) window mechanism of TCP. The AIMD algorithm controls the window size in response to the random marks from the bottleneck ECN/RED gateway. This model is introduced in Chapter 4, and we later refer to it as the *window-based model*.

While the models in Chapter 3 and 4 shed some insights on the interaction between the AIMD algorithm and RED, they do not represent a realistic depiction of the actual Internet because they ignore the session-level dynamics, *i.e.,* arrivals and terminations of flows, and the variability of round-trip delays. These additional layers of dynamics will be incorporated into the model in Chapter 5.

These models have common characteristics when taking the limit as the number of flows grows to infinity. First, properly normalized aggregated quantities (*e.g.,* queue size and throughput) converge, and simplified recursive dynamics emerge for the limiting quantities. Second, it is reasonable to expect the throughput of a flow to be asymptotically independent of other flows since the effect of any single flow on other flows becomes less pronounced as the number of flows grows large.

Here we identify the necessary ingredients for applying a limiting process to TCP model as done in this thesis: First of all, only the congestion in a bottleneck link is considered with capacity of the bottleneck link and the feedback information rate (from AQM) scaling with the number of users. For example, if

C denotes the capacity per user of at the bottleneck router, then its overall capacity should scale up to NC when N TCP flows share the bottleneck link. Also, let $f^{(N)}(x) : \mathbb{R}_+ \to [0,1]$ denote the feedback dropping/marking probability function in RED when the buffer level is x and there are N users in the system, *i.e.,* any incoming packets when the buffer level is x is randomly marked/dropped with probability $f^{(N)}(x)$. Then $f^{(N)}(x)$ should scale with N in the sense that there exists a continuous function $f : \mathbb{R}_+ \to [0,1]$ such that for each $N = 1, 2, \ldots,$

$$f^{(N)}(x) = f(N^{-1}x), \quad x \geq 0. \tag{1.1}$$

We are interested in the "snapshot" of the dynamics when there exists N flows in the system. Therefore, f is just a surrogate function representing the average contribution that each flow has on the dropping/marking probability.

Next, in order to describe the interaction between the bottleneck router and TCP flows, all models use a recursion similar to Lindley's recursion for the quantity of interest. In our study, this quantity is the queue size at the RED buffer. Let $Q^{(N)}(t)$ denote the queue size at the discrete time epoch t when the system has N flows. During the timeslot $[t, t+1)$, it is possible for queue size to increase (due to new incoming packets) or decrease (because the packets in the queue are serviced/transmitted). We denote the amount of increase and decrease in epoch t as $A^{(N)}(t)$ and $C^{(N)}(t)$, respectively. As the queue is always non-negative, we can write the following recursion

$$Q^{(N)}(t+1) = \left[Q^{(N)}(t) + A^{(N)}(t) - C^{(N)}(t)\right]^+. \tag{1.2}$$

Both $A^{(N)}(t)$ and $C^{(N)}(t)$ follow some stochastic recursions which depend on the current queue size and thereby forming a stochastic feedback system. For example, when $Q^{(N)}(t)$ is large, $A^{(N)}(t') - C^{(N)}(t')$ should be negative with high

probability for some $t' \geq t$ in order to drive the queue size down.

If we normalize (1.2) by N, the resulting recursion becomes

$$\frac{Q^{(N)}(t+1)}{N} = \left[\frac{Q^{(N)}(t)}{N} + \frac{A^{(N)}(t)}{N} - \frac{C^{(N)}(t)}{N}\right]^+ . \tag{1.3}$$

If $\frac{Q^{(N)}(t)}{N}$, $\frac{A^{(N)}(t)}{N}$ and $\frac{C^{(N)}(t)}{N}$ all converge in the same sense, *e.g.,* in probability, as N grows large to random variables (rvs) $q(t), a(t)$ and $c(t)$, respectively, then $\frac{Q^{(N)}(t+1)}{N}$ also converges, say to a rv $q(t+1)$, *i.e.,*

$$\frac{Q^{(N)}(t+1)}{N} \quad \stackrel{P}{\rightarrow}_N \quad q(t+1) \tag{1.4}$$

with

$$q(t+1) = [q(t) + a(t) - c(t)]^+ , \tag{1.5}$$

where $\stackrel{P}{\rightarrow}_N$ denotes convergence in probability when N tends to infinity. This produces a recursion for the rvs $\{q(t),\ t = 0, 1, \ldots\}$ with initial condition $q(0)$ and driving sequence $\{a(t), c(t),\ t = 0, 1, \ldots\}$. The model simplification typically results from the fact that the limiting rvs $a(t)$ and $c(t)$ are either deterministic quantities or rvs which are easy to evaluate, while in the model with arbitrary number of flows N, we typically need N or more state variables to evaluate either $A^{(N)}(t)$ or $C^{(N)}(t)$. The original quantity $Q^{(N)}(t)$ can be estimated by $N \cdot q(t)$. Since $q(t)$ can be evaluated independently of N, the estimation can be done efficiently, and the limiting model is scalable.

To provide a sharper approximation of $Q^{(N)}(t)$, we can seek to establish the following convergence

$$\sqrt{N}\left(\frac{Q^{(N)}(t)}{N} - q(t)\right) \Longrightarrow_N L_0(t) \tag{1.6}$$

for some rv $L_0(t)$ where convergence in distribution is denoted by $\Longrightarrow_N$ when N tends to infinity. If there exists such a rv $L_0(t)$, then (1.6) yields the (distributional) approximation

$$Q^{(N)}(t) \approx N \cdot q(t) + \sqrt{N} L_0(t) \tag{1.7}$$

in some distributional sense. Again, the rv $L_0(t)$ can be evaluated independently of N. We can then investigate the process $\{q(t), L_0(t),\ t = 0, 1, \ldots\}$ to derive both *qualitative* and *quantitative* properties concerning of the interaction between TCP flows and AQM mechanisms. The main theme of this thesis therefore revolves around establishing convergence results similar to (1.4) and (1.6), and investigating the properties of the limiting recursions.

1.3 Contributions

In this section, we highlight the main contributions of this thesis. The results are derived from various models with common characteristics described in the previous section. Indeed, the recursion of the RED buffer follows (1.2) with $A^{(N)}(t)$ and $C^{(N)}(t)$ left to be specified by the models. In all of the models, $C^{(N)}(t)$ represents the service capacity (in packets) of the bottleneck RED gateway which satisfies the scaling assumption outlined earlier, *i.e.*, $C^{(N)}(t) = NC,\ t = 0, 1, \ldots$, for some positive constant C where C denote the average capacity per flow. Therefore, the main differences between each model are in specifying the packet arrivals of the TCP sources in each timeslot.

In a distributed congestion-control scheme such as TCP, each source has to adjust its transmission rate by relying only on the local information at the source. This local information can be the state variables of the TCP

congestion-control mechanism and the past events (such as packet losses) that the TCP source has experienced. If $A_i^{(N)}(t)$ denotes the number of packets TCP source $i = 1, \ldots, N$ injects into the bottleneck gateway in timeslot $[t, t+1)$, then the aggregate number of packets coming into the RED buffer in timeslot $[t, t+1)$, $A^{(N)}(t)$ is given by

$$A^{(N)}(t) = \sum_{i=1}^{N} A_i^{(N)}(t). \tag{1.8}$$

Therefore, $Q^{(N)}(t) + A^{(N)}(t)$ packets are available for transmission during that timeslot. Since the outgoing link operates at the rate of NC packets/timeslot, $\left[Q^{(N)}(t) + A^{(N)}(t) - NC\right]^+$ packets remains during timeslot $[t, t+1]$ in the buffer, their transmission being deferred to subsequent timeslots. The number $Q^{(N)}(t+1)$ of packets in the buffer at the beginning of timeslot $[t+1, t+2)$ is therefore given by

$$Q^{(N)}(t+1) = \left[Q^{(N)}(t) - NC + A^{(N)}(t)\right]^+. \tag{1.9}$$

The evolution of $A_i^{(N)}(t)$ depends on $Q^{(N)}(t)$ through some signaling mechanisms, *e.g.,* random marking/dropping of packets with probability being a function of $Q^{(N)}(t)$. In all the models, we assume the structural condition (1.1) on how the marking/dropping probability function scales with N.

Under the aforementioned framework, the main theoretical contributions to be reported here are:

The Weak Laws of Large Numbers (WLLN)

Under different models, Theorems 1, 2, and 4 show that the dynamics of the RED buffer at time t, denoted as $Q^{(N)}(t)$, can be approximated by $Nq(t)$ with

$q(t)$ determined via a simple deterministic recursion, which is independent of the number of users. The limiting model is therefore "scalable" as it does not suffer from the state space explosion, nor does it require any ad-hoc assumptions. Indeed, we have the convergence (1.4), so that the approximation becomes more accurate as the number of flows becomes large. We also note that (1.4) is a byproduct of the convergence

$$\frac{\sum_{i=1}^{N} A_i^{(N)}(t)}{N} \xrightarrow{P}_N a(t) = \mathbf{E}\left[A(t)\right], \tag{1.10}$$

for some constant $a(t)$ and rv $A(t)$ determined by the model. The convergence (1.10) can be interpreted as a Weak Law of Large Numbers for the array $\{A_i^{(N)}(t),\ N = 1, 2, \ldots;\ i = 1, \ldots, N\}$.

Moreover, the dependency between each TCP flow with RED becomes negligible under a large number of flows. This finding is compatible with the belief that RED breaks the global synchronization between TCP flows [18].

These Weak Laws of Large Numbers also provide some justification for evaluating the dynamics of TCP/RED (and more generally TCP/AQM) through deterministic models as surveyed in Chapter 2. The Weak Laws of Large Numbers suggest that the recursion of the average queue $q(t)$ depends only on the average workload input to the network $a(t) = \mathbf{E}\left[A(t)\right]$ and the distributional recursion of $A(t)$ is closely related to the recursion of a single flow.

Steady-state analyses

For N persistent TCP flows when the bottleneck capacity is NC, in non-trivial situations, the average (steady-state) throughput of each flow is seen to be approximately C for large N, so that TCP traffic does not realize the benefits

commonly associated with statistical multiplexing. Indeed, for N open-loop independent traffic sources, each operating at average rate C, the bandwidth required to serve the aggregate flow is typically less than NC under statistical multiplexing. TCP flows, on the other hand, are correlated due to their coupling via the bottleneck router, and multiplexing TCP flows is not as effective.

To make use of the limiting model in network dimensioning, the queue length at the steady-state can be easily calculated, and the steady-state distribution of the window size can be evaluated from well-known TCP models [47, 48]. For TCP flows with session dynamics and variable round-trip delays, we propose estimating the steady-state limiting queue size by a simple approximation under mild assumptions. It is noteworthy that the average buffer utilization and average window size at the steady-state depend on the RTT and on the file size distributions only through their mean values.

The Central Limit Theorems (CLT)

For a more accurate description of the queue dynamics, Theorems 3 and 5 present a Central Limit analysis validating the convergence (1.6) for some rv $L_0(t)$, provided the marking/dropping probability function is continuously differentiable. The rvs $\{L_0(t),\ t = 0, 1, \ldots\}$ are Gaussian except in a special technical case and a distributional recursion is also provided for generating them.

The Central Limit analysis leading to (1.6) reveals that the magnitude of the queue fluctuation is proportional to the derivative of f around the limiting (normalized) queue size $q(t)$. This advocates the use of a smooth marking probability function as suggested in the RED "gentle" option as opposed to the original recommendation in [16] and [18]. This finding coincides with the

observed oscillatory behavior of RED when the average packet drop rate exceeds max_p, in the absence of RED's "gentle" modification [14].

A closer inspection at the CLT results reveals the sources of queue fluctuations can be decomposed into

(a) The fluctuation in the feedback information, *i.e.,* the difference between the limiting feedback probability $f(q(t))$ and the actual feedback probability $f^{(N)}(Q^{(N)}(t))$.

(b) The binary nature of the feedback information, *i.e.,* the sources have to approximate the feedback probability $f^{(N)}(Q^{(N)}(t))$ from whether the transmitted packets received any marks.

(c) The variation in the RTTs and file size variations.

The limiting rvs representing (b)-(c) are Gaussian and their combined effects on the queue fluctuations are determined through the protocol structure.

1.4 Thesis Organization

This thesis is organized as follows: In Chapter 2, we begin with a literature review that relates to our work. In Chapter 3, we describe a simple ersatz model of homogeneous, persistent rate-controlled flows which react to the events in the network, *i.e.,* congestion/no congestion. This ersatz model illustrates the type of results than can be expected in more complex situations. Chapter 4 presents a more realistic model of persistent TCP flows which incorporates the explicit AIMD window mechanism of TCP reacting to the marking mechanism in the RED bottleneck gateway. The Weak Law of Large Numbers of the model is

presented in Chapter 4 along with its Central Limit analysis. In Chapter 5, the model is then extended to incorporate session-layer dynamics and variable round-trip delays — the Weak Law of Large Numbers and a Central Limit analysis are established for this model. The proofs of the Weak Laws of Large Numbers are given in Chapter 6 while the proof of the Central Limit theorem are given in Chapter 7. Finally in Chapter 8, we summarize the work that has been done in the dissertation and discuss future research directions.

1.5 Notation

Some words on the notation in use: 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 $\xrightarrow{P}_n$ (resp. $\Longrightarrow_n$) to denote convergence in probability (resp. weak convergence or convergence in distribution) with n going to infinity. For scalars a and b we write $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For a fixed positive integer N, let $\mathcal{N} = \{1, \cdots, N\}$ denote the set of sessions and let $X^{(N)}$ indicate the explicit dependence of the quantity X on the number N. The expectation of a rv X with a distribution function F is given by either $\mathbf{E}[X]$ or $\mathbf{E}[F]$. For simplicity, we introduce the notation $\mathbf{1}_X[x]$ and $\mathbf{P}_X[x]$ for $\mathbf{1}[X = x]$ and $\mathbf{P}[X = x]$, respectively.

Chapter 2

Literature Survey

This chapter discusses both previous research and the current state-of-the-art relating to our work. The chapter is organized as follows. First we present an overview on several congestion-control mechanisms and active queue management (AQM) algorithms in Section 2.1. Next, we survey the work on the modeling of these congestion-control mechanisms and their interaction with AQM mechanisms. We categorize different modeling approaches into two classes. The first class consists of models derived without the assistance of limit theorems. Models in this class will be referred later as *micro-scale* models and are presented in Section 2.2. The second class consists of models derived via limit theorems. We later refer to models in this class as *macro-scale* models and present them in Section 2.3. Models which will be developed in this thesis also belong in this class.

2.1 Overview of Congestion-Control and AQM Mechanisms

Two major classes of congestion-control mechanisms are surveyed: (i) TCP congestion-control mechanism and its variants, and (ii) Optimization-based

congestion-control mechanisms. For AQM mechanisms, we first present the earliest and most widely-implemented algorithm, Random Early Detection (RED). Other popular AQM mechanisms are briefly described later.

The Internet and TCP congestion-control

The Internet was developed by the United States Department of Defense prior to and concurrent with ISO standard activities. The resulting internetwork, known as ARPANET, has now been extended to incorporate internetworks developed by other government agencies, academic institutions and commercial networks . The combined internetworks are now known simply as the Internet [20].

The protocol suite used within the Internet is known as the Transmission Control Protocol/Internet Protocol (TCP/IP) suite. Since all the protocol specifications associated with TCP/IP are available in the public domain, they have been used extensively by commercial and public organizations for creating open system networking environments.

The Internet Protocol (IP) main functions are equivalent to those of the OSI networking layer. These functions include fragmentation and reassembly, routing, and error reporting. All are necessary for internetworking across dissimilar networks [20].

The Transmission Control Protocol (TCP) is a connection-oriented transport protocol. Its service offered to the users is known as "reliable stream transport service". Its functions cover (but are not equivalent to) those of the OSI Transport layer and Session layer. Besides establishing and terminating connections, TCP is also needed for end-to-end control over the "best-effort" service of IP. The role of TCP, which is hidden from users, is to adapt the

sending rate of the source to the capacity of the network. Without such a *congestion-control* mechanism in TCP, traffic sources could saturate one or several routers, thereby causing many losses and retransmissions and resulting in congestion collapse [26].

At the core of TCP is a dynamic window-based congestion-control scheme introduced by Jacobson [26]. The window mechanism limits the number of packets sent by the source and not yet acknowledged by the destination. This limit on the number of unacknowledged packets is referred as *the congestion window size*. The efficiency of the congestion-control mechanism depends greatly on the choice of the window size. If the window size is too large, the source could saturate the routers. If the window size is too small, the link may not be fully utilized. Due to the complex nature of the Internet, the optimal value of the congestion window size is not known *a priori* because it depends on dynamic parameters of the network such as the number of connections sharing the same links. Moreover, the optimal window size can also change during the duration of a connection because of the time-varying nature of the Internet. As a result, the window size of TCP adapts according to the network conditions.

There are two working phases in the TCP congestion-control mechanism. The first phase is called the *slow start* phase where the window size is increased by one packet for every acknowledgment received. Therefore, the congestion window size doubles approximately every round-trip time[1], resulting in an exponential increase of the window size as a function of time. The second phase is called the *congestion avoidance* phase, during which the window size increases linearly as a

[1]Round-trip time (RTT) is the duration of time since the beginning of the transmission of a packet to the destination until an acknowledgment from the recipient is correctly received.

function of time. TCP initially operates in the slow start phase trying to quickly gauge the available bandwidth. When a packet loss is detected, TCP cuts back its window size trying to resolve the congestion in network, and then switches to the congestion avoidance mode. The actual implementation of these mechanisms varies with each version of TCP.

The most popular version of TCP, called *Reno*, employs an *additive increase, multiplicative decrease (AIMD)* algorithm [26]. In this algorithm, the TCP congestion window size will increase by approximately one packet per round-trip in the congestion avoidance phase. When congestion is detected by packet loss, the window size is reduced by half. As a result, the TCP Reno congestion window oscillates between periods of under-utilization and over-utilization of network resources. Under certain conditions, it can be shown that TCP Reno flows with the same round-trip delay competing for the bandwidth in a single bottleneck have equal long-term average throughput [7].

In order to avoid such an oscillatory behavior, a new version of TCP, called *Vegas*, was proposed in [9]. The main idea is to use the delay information of the packets instead of packet losses in order to adjust the congestion window size. TCP Vegas can stabilize the window size close to the optimal value without introducing periodic oscillations. However, the incompatibility between TCP Reno and Vegas is well-documented [44] and presents a major obstacle in the deployment of TCP Vegas. Therefore, it is reasonable to expect that TCP Reno will remain the dominant congestion-control algorithm in the Internet for the foreseeable future.

Optimization-based congestion-control

One approach which has received much attention recently is to model TCP throughput as the solution to a utility maximization problem. The interest in this model originates from the work of Kelly [31]. It shows that the utility maximization problem of the system composed of a network and of users can be decomposed into two separate problems, namely the network problem and the users problem, assuming the utility function of each user is a concave function of the user's received throughput. If the network maximizes its revenue and then users subsequently maximize their utilities, the recursive maximization sequences will converge to the solution of the system utility maximization problem. Subsequent work by Kelly et al. [30] shows that the user problem can be solved by a rate control algorithm which can be implemented as a congestion-control algorithm. The rate control algorithm depends on the form of the utility function of the user. Recent work by Low [36] shows that the throughput of the AIMD TCP congestion-control algorithm solves the user problem with a specific utility function. Furthermore, AQM mechanisms such as RED can be modeled as the feedback functions from the network to the users. This enables efficient modeling of a large network with multiple users and makes it possible to predict the interactions between the network and traffic flows through the choices of utility functions and network parameters.

Random Early Detection (RED) gateway

Random Early Detection (RED) is one of the earliest AQM mechanisms, and is currently the most widely-implemented. It was proposed by Floyd and Jacobson [18] to keep the queue size small, reduce burstiness and solve the problem of

global synchronization in the Internet, *i.e.,* when several TCP flows compete for bandwidth in a tail-drop gateway, packets from many flows are usually discarded (near) simultaneously [63], resulting in a poor utilization of the network.

To accomplish these goals, RED signals incoming flows of incipient congestion by randomly dropping incoming packets with a probability depending on the exponentially averaged queue length. As a result, flows are notified and have enough time to react to growing congestion before the actual queue is full. RED is designed to work in conjunction with congestion-controlled TCP flows. RED also incorporates a marking mechanism to signal flows of congestion through the Explicit Congestion Notification [15] option in the IP packet header.

The introduction of RED initiated a whole new research area in active queue management. There are now many proposed variations of RED and other similar AQM schemes, some of which will be surveyed in the next subsection. However, the interaction between AQM schemes and Internet traffic even for a simple AQM scheme such as RED is still not well understood as there exist several conflicting reports. For example, while it is claimed that RED can avoid global synchronization in [18], it is reported in [8] that the number of consecutive packet drops in RED is higher than in tail-drop gateway, and therefore that RED does not help alleviate the global synchronization problem.

Other active queue management mechanisms

This section provides a brief overview on some other interesting AQM mechanisms. Although all of these schemes mark or drop packets to signal flows of incipient congestion similarly to RED, they have more complex mechanisms to determine the feedback probability.

The PI controller [24] is proposed as a follow-up to the work in [22] that models the TCP/AQM interaction in the framework of a linearized feedback control system. The proposed PI controller improves the transient behavior and forces the queue to achieve the target buffer utilization in the steady-state.

BLUE [13] is an active queue management mechanism which uses packet losses and link idle events to adjust the packet dropping probability instead of the average queue size in RED. By decoupling the queue length from congestion management, the authors claim that BLUE can reduce the packet loss rate in the queue and requires smaller queue in order to achieve the stable operating regime.

Stabilized RED [46] uses a packet sampling technique to estimate the number of flows utilizing the bottleneck link. The appropriate dropping probability in the bottleneck link is then computed on the basis of this estimate, since the number of flows is an indicator of the fraction of the queue which would be reduced by triggering the multiplicative decrease mechanism in TCP congestion-control.

A modification to RED called Adaptive RED is proposed in [17]. In this scheme a target queue utilization is selected and the marking/dropping probability is adjusted by an additive increase/multiplicative decrease mechanism on every predefined interval. If the queue is smaller than the target queue, the marking probability is decreased multiplicatively, otherwise it is increased by some small constant. It is shown in [35] that Adaptive RED is insensitive to the network load and induces smaller fluctuation than RED.

Kelly et al. introduces a virtual queue mechanism as an alternative to congestion management [31]. The virtual queue mechanism maintains a fictitious queue with a smaller capacity and buffer size than the actual queue. For every packet arrival or departure in the actual queue, the virtual queue is updated.

Whenever the virtual queue overflows, any packet arrival in the actual queue is marked/dropped to signal the sources of incipient congestion. The virtual queue mechanism requires smaller queue to operate than RED but cannot remedy the flow synchronization problem.

A problem with the virtual queue mechanism is that the appropriate capacity of the virtual queue depends on both the load and the round-trip delay of the flows. Kunniyur and Srikant suggest an improved virtual queue mechanism called Adaptive Virtual Queue [34], in which the capacity of the virtual queue is adjusted through a slower control loop. It is claimed that the Adaptive Virtual Queue mechanism yields very high utilization, very small packet loss and queueing delay for a wide-range of traffic load.

A very desirable property for AQM mechanisms is to penalize or regulate aggressive/unresponsive flows. CHOKe [51] is a very simple mechanism operating on top of the original RED algorithm. It can be shown that the simple, stateless CHOKe algorithm can guarantee a bound on the throughput of unresponsive flows.

It is easy to imagine that the additional complexities in these AQM mechanisms further complicate the analysis of the TCP/AQM interaction. A systematic evaluation of different AQM mechanisms is needed to verify and compare the claims in each study. It is our hope that the framework proposed in this thesis can be generalized to encompass these AQM mechanisms.

2.2 Micro-scale Modeling of Congestion-Controlled Flows

In this section, we review some of the existing work on micro-scale traffic modeling of a congestion-controlled traffic flow as well as the efforts made to extend these micro-scale models to the situation with many interacting flows.

TCP modeling by Markov processes

One of the earliest and most popular efforts to model TCP made use of Markov chain modeling. The size of the TCP congestion window acts as the state of the chain and the loss probability (either independent or dependent on the state) determines the transition probabilities. Padhye et al. have derived an approximation for the steady-state distribution of a Markov chain modeling a single TCP connection with a fixed loss probability [48]. Altman et al. [2] extended the model to cover more general loss processes, *e.g.,* continuous-time Markov chain with different loss probability in each state.

Extension of micro-scale TCP model to many TCP flows

In this section, we point to various efforts for extending the micro-scale TCP model to many TCP flows. In the models to be described, there is no attempt to consider a limiting model.

The resulting models suffer from either one of the following difficulties: (i) These models are typically too complex and therefore not tractable analytically. Sometimes this can also translate in numerical calculations being computationally prohibitive; (ii) Reduction in complexity can be achieved by making simplifying assumptions which are often *ad-hoc*, *difficult to justify*, and

sometimes downright *unrealistic*. We argue that these assumptions should emerge in a natural way from the traffic modeling process, rather than being enforced for the main reason that they enable the analysis to go through. It will soon become apparent that a modeling methodology based on limit theorems provides a direct remedy to this issue – Indeed, model simplifications occur in the limit, without the need for any ad-hoc assumptions.

Hasegawa et al. [21] consider a Markov chain for N TCP connections using either Tail-Drop or RED gateway. The state space in this system is the vector of window size of all of N connections and the queue size, which is a function of the sum of the size of congestion windows. For each connection, the transition probability depends on the current window size and queue size, hence on the window size of *all* connections. The authors then derive a fixed-point solution of the average window size in steady-state. However, this fixed-point solution is very complicated and requires solving a large system of non-linear equations. It is not clear how this could be accomplished effectively and whether it would offer any analytical understanding to the problem.

Another example is the work by Garetto et al. [19] where a closed-queueing network model is proposed to investigate the interaction between many TCP flows. A closed network of $M|M|\infty$ queues is introduced where each queue represents a state of the TCP algorithm. The number of users in each queue represents the number of TCP connections in that state. The service rate of the user in each queue depends on the TCP state. After completing service in a queue, each user's transition to the next queue depends on the loss probability. This closed-queueing network is used in a two-tier model which can be described as follows: Given a loss probability, a numerical calculation of the steady-state

distribution of the closed-queueing network yields an approximated traffic load. This load will be input into a pseudo-network model (*e.g.,* $M|M|1|B$) to determine the new loss probability for the closed-queueing network. If the result from the successive iterations converges, it is claimed that the convergence will be to the same working point as given by the model where each TCP flow is modeled in details.

Several questions arise concerning such a model. First, the model implicitly assumes that the dynamics of the TCP congestion-control mechanism converge to steady state faster than the network dynamics; this is the opposite of the real Internet where TCP reacts to events in the network. Next, while the numerical calculations are not as complex as would be the case in a detailed model of TCP flows, they are far from simple. The example considered in [19] consists of 11 different queue types and the number of queues (M_q) is 357. Calculations for the steady-state distribution of the queue are typically of the order $O(M_q^2)$ for each iteration. Finally, the "interaction" between TCP flows is developed in an abstract model where its deviation from the actual interaction process cannot be quantified. It also impossible to draw any analytical conclusion concerning these interactions.

Baccelli et al. [4] propose fixed-point methods for the simulation of the sharing of a local loop by a large number of interacting homogeneous TCP connections. The analysis uses a detailed description of one TCP connection and a simplified description of the interaction with other connections. It is again difficult to quantify the effect of such simplifications and the accuracy of the results can only be verified by extensive simulation.

Fluid approximation of TCP congestion-control mechanism

Mathis et al. [40] use a continuous-time fluid flow approximation to the discrete time process describing window behavior. Assuming the congestion signals TCP to back off according to a Poisson process, where the k^{th} signal occurs at time epoch τ_k $(k = 1, 2, \ldots)$, then the approximate evolution of the congestion window evolution of a single TCP flow is governed by

$$\frac{dW(t)}{dt} = \frac{1}{W(t)}, \tag{2.1}$$

except at the points τ_k $(k = 1, 2, \ldots)$ where

$$W(\tau_k^+) = W(\tau_k^-)/2. \tag{2.2}$$

These equations can be used to derive the fact that that the average window size is of the order of $1/\sqrt{p}$. This model is suitable only for a single flow because the congestion notification is assumed to be Poisson and each notification is independent of each other (an assumption similar to [43] which uses stochastic differential equation to model TCP). When more than one flow utilizes the same bottleneck link, this assumption is not helpful for capturing the interaction between flows. Bonald [7] considers a similar model for several TCP flows with the major difference that congestion occurs when the sum of congestion windows exceeds the bandwidth-delay product plus the bottleneck buffer. Under the assumption that at every congestion epoch *all* TCP flows simultaneously back off, TCP can be shown to be fair and an explicit closed-form formula for the utilization can be derived. However, the assumption of total synchronization between flows is unrealistic.

While there are many advantages of viewing the system as a utility maximization problem and of modeling TCP traffic as rate-controlled fluid flows,

there are definite drawbacks as well. Most important of all is the absence of "packets" from the system model, although the congestion-control mechanism of TCP relies on packet-level operations. For example, it is not possible under this model to derive the queue length distribution at routers which is an important quantity for any successful network dimensioning. Additionally, while the solution to the maximization problem might accurately describe the steady-state solution of TCP, the distributed solution does not necessarily capture the short-term dynamics of TCP well. Finally, the numerical calculation of the solution is still very complex as it suffers from state space explosion when the number of TCP flows becomes large. This type of models appears more suitable for understanding the "big picture" and the qualitative behavior of congestion-control algorithms, rather than for effective and accurate network dimensioning.

Stability of the congestion-control and AQM mechanisms

As previously mentioned, the interaction between the congestion-control and AQM mechanisms can be analyzed as feedback control systems. Using the optimization-based framework, Kelly et al. have shown the stability of the system in the absence of delay through a Lyapunov function argument [30]. Subsequently, a lot of results have been published on both the local and global stability of the system with arbitrary delay either in the context of the classical TCP congestion-control algorithm or of utility-based congestion controllers (*e.g.,* [28, 38, 49, 54]).

While the aforementioned work yields great insights into the dynamics of the control system that governs the congestion-control/AQM interaction, questions

remain. First of all, these results are based on deterministic models of feedback-delay systems. However, the relationship between the actual systems (which are stochastic) and the deterministic models are not apparent. For example, what is the relationship of the probabilistic marking mechanism in AQM implementation to the shadow price in the utility-based model? Furthermore, the stability conditions often assume a small number of homogeneous flows in order to obtain analytically tractable results due to the complex dynamics and large state-space involved. Sometimes, the stability conditions are stated in the "worst-case" conditions, *e.g.,* the largest delay, which could be extremely conservative.

This thesis remedies to these shortcomings as follows. First, it provides a justification on using the deterministic models to analyze the stability of the systems. More specifically, in the large number of flows regime, the dynamics of the average queue depends only on the dynamics of the average traffic flows in/out of the queue. The dynamics of the average traffic are also closely related to the dynamics of a single flow in the case of homogeneous flows, lending support to the practice of considering the simple deterministic models. For systems with heterogeneous round-trip delays, we also derive a simple recursion for the average queue and traffic flow. By analogy with the homogeneous round-trip case, we expect that a simple deterministic model should also exist for the analysis of the stability of the system. This would relax the stability conditions as we do not have to rely on the worst case conditions.

Modeling of TCP traffic with session-layer dynamics

Existing literature on short-lived TCP traffic modeling usually relies on ad-hoc assumptions, which ensure the model accuracy only in certain regimes. Hollot et al. model short-lived TCP flows as exponential pulses, *i.e.,* time-shifted, increasing exponential functions of limited durations, whose interarrival times are exponentially distributed, *i.e.,* Poisson process [23]. The statistics of these exponential pulses can be characterized through a time-reversal of a well-known class of processes called *shot noise processes*. This model assumes that the short-lived flows last only a few round-trip times and do not experience packet drops or marks, thereby implicitly assuming that congestion level is relatively low. Furthermore, flows are always in either congestion avoidance (long-lived connections) or slow start (short-lived connections), and do not transition from one to the other. In other words, the session dynamics, where connections arrive and leave the network after transfers are completed, are not explicitly modeled. A similar approach to modeling short-lived flows is also taken in [42].

On the other end of the spectrum, Kherani and Kumar suggest that as the bottleneck capacity becomes very small, the queueing model for the bottleneck queue can be accurately described as a processor sharing queue [32]. When the capacity is large, however, the processor sharing model becomes less accurate because newly arrived TCP flows cannot fully utilize their allocated bandwidth. In fact, in the large capacity regime these short-lived flows may terminate even before they can increase their transmission rates to fully utilize their allocated bandwidth due to slow start.

In Chapter 5, we analyze a model which explicitly incorporates the session-layer dynamics and show that the limiting model agrees with these

models in their respective regime, *i.e.,* when the capacity is either very large or very small.

Modeling of TCP traffic with heterogeneous round-trip delays

Little work has been done on modeling the interaction between an AQM mechanism and a large number of TCP flows with diverse delays. Some initial investigation of the role of heterogeneous RTTs into such an interaction is presented in [38]. 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 create 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.

The role of the variable round-trip delays to the interaction between TCP and AQM interaction in the large number of flows regime is investigated in details in Chapter 5.

2.3 Macro-scale Modeling of Congestion-Controlled Flows

In this section, we outline the type of results that can be expected from modeling TCP with a large number of homogeneous TCP flows via limit theorems.

All of the existing models surveyed in this section ignore the packet-level operations of TCP, and consider only the evolution of the transmission rate, while models in this dissertation take into account detailed packet-level

operations. It is important to make this distinction because the actual operations of TCP, of RED and of the network rely on the packet-level operations. The omission of this level of details simplifies the analysis but at the expense of additional distortion in the results. For example, it will be demonstrated in Theorem 5 that the ECN marking in packets introduce additional fluctuation in the limiting queue. This fluctuation cannot be captured through the models without packet-level operations.

Throughout this section, we use a common notation $x_i^{(N)}(t)$ to denote the transmission rate of flow i $(i = 1, \ldots, N)$ at time t.

Shakkottai and Srikant [55] consider a discrete-time model of N homogeneous proportional-fair, congestion-controlled flows (such as the primal algorithm in [30]). These flows utilize a bottleneck router with either rate-based marking (such as Virtual Queue marking) or queue-based marking (such as RED), *i.e.,*

$$x_i^{(N)}(t+1) = \left(x_i^{(N)}(t) + \Delta - \beta x_i^{(N)}(t-d) f^{(N)}(x_i^{(N)}(t-d) + e_i^{(N)}(t-d)) \right)^+,$$

where Δ and β are positive constants which determine the rate at which a flow increases and decreases its transmission rate, f is the marking function, d is the round-trip delay between the flow and the bottleneck router, and $e_i^{(N)}$ is a "noise" process, representing short-lived and uncontrolled flows. A natural way to obtain a simplified limiting model resorts to rescaling the length of timeslots to be inversely proportional to the number of flows N. Then, as N gets large, the average transmission rate is expected to converge almost surely to a deterministic quantity (under appropriate assumptions on the "noise" process and the function f) . This quantity can be described by a functional differential equation, thereby justifying deterministic fluid approximations of the algorithms described in Section 2.2. However, the aforementioned approach does not apply well to the

Internet with TCP congestion-control because (i) the model ignores the complexity of the window-based implementation and considers only the transmission rate of the flow, and (ii) the congestion-control algorithms considered are also derived from the solution to utility maximization problems (as described in Section 2.1), whence continuous in their fluid limit and thus "nicer" than TCP congestion-control algorithm which may experience abrupt transmission rate changes. It is still an open technical problem on whether there exists a fluid limit for AIMD TCP congestion-control.

Hong and Lebedev [25] consider a model where the throughput of each connection evolves at time epochs when congestion occurs. At every epoch, each connection draws a $\{0.5, 1\}$ random number representing the fraction of throughput that the connection retains after congestion. This fraction depends on the throughput of the flow just before congestion occurred. If T_n is the n^{th} congestion epoch and $\gamma_{i,n}(x)$ is a $\{0.5, 1\}$-valued rv depending on the rate x, then

$$x_i^{(N)}(T_n^+) = \gamma_{i,n}\left(x_i^{(N)}(T_n^-)\right) \cdot x_i^{(N)}(T_n^-)$$

with an obvious notation. Let NC be the capacity of the bottleneck router, and assume that the residual capacity at time T_n^+, namely $NC - \sum_{i=1}^N x_i^{(N)}(T_n^+)$, is divided evenly among all users. Then, [2]

$$x_i^{(N)}(T_{n+1}^-) = x_i^{(N)}(T_n^+) + C - \tfrac{1}{N}\sum_{i=1}^N x_i^{(N)}(T_n^+).$$

Denote $x_{i,n}^{(N)}$ the transmission rate of connection i right after the n^{th}-epoch.

[2]The authors' model is flawed since the transmission rate in the model can be negative. However, the extension of the model to incorporate a restriction on the rate to be only positive should not change the nature of the result.

Under certain conditions on the rv $\gamma_{i,n}$, the following limits exist:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_{i,n}^{(N)} = s_n \quad \text{in } L_1 \text{ and a.s.},$$

where $s_n = \mathbf{E}\left[x_n^{(\infty)}\right] = \lim_{N\to\infty} \mathbf{E}\left[x_{i,n}^{(N)}\right]$. Moreover, the sequence $\{x_n^{(\infty)},\ n = 0, 1, \ldots\}$ is stationary and ergodic, and can be characterized by the relation

$$x_{n+1}^{(\infty)} \quad =_{st} \quad \gamma_{n+1}\left(x_n^{(\infty)} + C - \mathbf{E}\left[x_n^{(\infty)}\right]\right) \cdot \left(x_n^{(\infty)} + C - \mathbf{E}\left[x_n^{(\infty)}\right]\right),$$

where $\gamma_{n+1}(x)$ is a rv which has the same distribution as $\gamma_{i,n+1}(x)$.

This result suggests the existence of a simpler process for describing the asymptotic behavior of the average rate. Although the model unrealistically assumes that the multiplicative decrease in a TCP flow depends only on its own transmission rate and not on that of other flows, we will show in the next section that similar results also exist in more elaborate models with many TCP flows sharing a RED gateway.

Adjih et al. [1] consider a continuous-time model where partial differential equations of the free buffer space in a Tail-Drop gateway and the density of the window size distribution are specified. Under the assumption that the buffer in the bottleneck router scales with the number of users, some asymptotic results on both the free buffer space and the window distribution are established by the use of mean-field approximations.

Baccelli et al. [5] consider a continuous-time model where the window evolution for each user is described by a stochastic differential equation. For each of the window, the rate grows linearly and inversely proportional to the round-trip delay. However, the window size will be cut in half at a random time determined by a modulated Poisson process. The modulated Poisson processes

for all flows are coupled through a common rate function which depends on the queue size. As the number of flows grow large, the average queue size and the window size can be determined through a deterministic mean-field system.

Chapter 3

A Rate-based Model for TCP/RED

In this chapter, we present a stochastic model that captures the essential features of TCP congestion-control, *i.e.,* the gradual increase and the sudden decrease of transmission rate, combining with a random drop algorithm similar to RED. We analyze this *ersatz* model as the number of competing flows becomes large, and show under very mild assumptions that the stochastic model simplifies to a two-dimensional deterministic recursion when the number of flows grows large. This result suggests that with a large number of flows, network operators might be able to easily estimate the aggregate behavior of TCP flows and to dimension network resources accordingly.

First, the ersatz model is presented in Section 3.1. The model is organized around the behavior of the RED buffer contents observed at discrete epochs. While the model presented in this chapter is very simple, it provides a complete conceptual picture of the problem, *i.e.,* the growing size of the state space as the number of flows increases. In Section 3.2, we present the limit theorem of this simple model which demonstrates the type of results to be expected in a more complicated and realistic model, *i.e.,* the simplification and the limiting recursion through the law of large numbers. Results from Monte-Carlo simulations of the

model demonstrating the asymptotic behavior are shown in Section 3.3.

3.1 The Model

Fix $N = 1, 2, \ldots$, and consider the situation where N flows are active. For simplicity time is assumed slotted. Let $Q^{(N)}(t)$ denote the number of packets in the buffer at the beginning of timeslot $[t, t+1)$. In that timeslot, source i generates $A_i^{(N)}(t)$ packets according to a mechanism to be specified shortly. Following the approach outlined in Section 1.3, we have the recursion of the queue $Q^{(N)}(t)$ as (1.9) with $A^{(N)}(t)$ given by (1.8).

In order to fully specify the model, we need to specify the *joint* statistics of the rvs $\{A_i^{(N)}(t),\ i = 1, \ldots, N;\ t = 0, 1, \ldots\}$. This will be done in details later. Throughout, let $f^{(N)} : \mathbb{R}_+ \to [0, 1]$ denote the *dropping probability* function of the RED gateway. Moreover, we find it convenient to use the collection of i.i.d. $[0, 1]$-*uniform* rvs $\{U_i(t+1), V_i(t+1),\ i = 1, \ldots;\ t = 0, 1, \ldots\}$ which are assumed independent of the rvs $Q^{(N)}(0)$ and other initial conditions.

In this model, a source either transmits or is idle in a given timeslot. So, let $B_i^{(N)}(t+1)$ be a $\{0, 1\}$-valued rv that encodes packet generation by source i. Moreover, let $R_i^{(N)}(t+1)$ represent the possibility that the packet generated by source i at the beginning of timeslot $[t, t+1)$ is rejected, *i.e.,* $R_i^{(N)}(t+1) = 1$ (resp. $R_i^{(N)}(t+1) = 0$) if the packet is rejected by (resp. accepted into) the RED buffer. Set

$$B_i^{(N)}(t+1) = \mathbf{1}\left[U_i(t+1) \leq \alpha_i^{(N)}(t)\right] \tag{3.1}$$

where $\alpha_i^{(N)}(t)$ is an $[0, 1]$-valued rv which denotes the (conditional) *transmission*

rate of traffic source i at the beginning of timeslot $[t, t+1)$, and let

$$R_i^{(N)}(t+1) = \mathbf{1}\left[V_i(t+1) \leq f^{(N)}(Q^{(N)}(t))\right] \tag{3.2}$$

denote the indicator function of the event that the incoming packet from source i will be rejected. Thus,

$$A_i^{(N)}(t) = (1 - R_i^{(N)}(t+1))B_i^{(N)}(t+1). \tag{3.3}$$

To select the transmission rates we argue as follows: Suppose that source i generates no packet during timeslot $[t, t+1)$ ($B_i^{(N)}(t+1) = 0$), then the transmission rate of source i in the next timeslot remains unchanged. If on the other hand, a packet is produced by source i at the beginning of timeslot $[t, t+1)$, then either the packet is successfully transmitted ($R_i^{(N)}(t+1) = 0$), or it is dropped ($R_i^{(N)}(t+1) = 1$). In the former case, the transmission rate of source i in the next timeslot is *increased* to $\alpha_i^{(N)}(t)^{1-\varepsilon}$ $(0 < \varepsilon < 1)$, while this transmission rate is *decreased* by a factor γ $(0 < \gamma < 1)$ to $\gamma\alpha_i^{(N)}(t)$ in the latter case.

Under the constraint that transmission rates are bounded to the unit interval, these rules attempt to *emulate* the additive increase and multiplicative decrease, respectively, of the TCP congestion-control by conservatively increasing the transmission rate if the transmission is successful and reducing the transmission rate by the factor γ in the event of a packet loss. This can be summarized into the single equation

$$\begin{aligned} & \alpha_i^{(N)}(t+1) \\ = \; & \alpha_i^{(N)}(t)^{1-\varepsilon}(1 - R_i^{(N)}(t+1))B_i^{(N)}(t+1) + \; \gamma\alpha_i^{(N)}(t)R_i^{(N)}(t+1)B_i^{(N)}(t+1) \\ & + \alpha_i^{(N)}(t)(1 - B_i^{(N)}(t+1)). \end{aligned} \tag{3.4}$$

3.2 The Weak Law of Large Numbers

We are interested in determining the limiting behavior of the rate-based model as the number of sources N becomes large. The discussion is carried out under the following assumptions (AR1)-(AR2):

(AR1) There exists a continuous function $f : \mathbb{R}_+ \to [0,1]$ such that (1.1) holds for each $N = 1, 2, \ldots$, *i.e.,*

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

(AR2) For each $N = 1, 2, \ldots$, the queue dynamics start with the conditions

$$Q^{(N)}(0) = 0 \quad \text{and} \quad \alpha_i^{(N)}(0) = \alpha, \quad i = 1, \ldots, N$$

for some non-random α in $(0, 1]$.

Assumption (AR1) is the structural condition discussed in Section 1.3, while (AR2) is made essentially for technical convenience as it implies that for each $N = 1, 2, \ldots$ and all $t = 0, 1, \ldots$, the rvs $\alpha_1^{(N)}(t), \ldots, \alpha_N^{(N)}(t)$ are *exchangeable.* Assumptions (AR2) can be omitted but at the expense of a more cumbersome discussion.

Theorem 1. *Assume (AR1)-(AR2) in the rate-based model of Section 3.1. Then, for each $t = 0, 1, \ldots$, there exist a (non-random) constant $q(t)$ and a $[0,1]$-valued rv $\alpha(t)$ such that the following holds:*

(i) The convergence

$$\frac{Q^{(N)}(t)}{N} \xrightarrow{P}_N q(t) \quad \text{and} \quad \alpha_1^{(N)}(t) \xrightarrow{P}_N \alpha(t)$$

takes place;

(ii) For any integer $I = 1, 2, \ldots$, the rvs $\{\alpha_i^{(N)}(t),\ i = 1, \ldots, I\}$ become asymptotically independent as N becomes large, with

$$\lim_{N\to\infty} \mathbf{P}\left[\alpha_i^{(N)}(t) \le x_i,\ i = 1, \ldots, I\right] \ = \ \prod_{i=1}^{I} \mathbf{P}\left[\alpha(t) \le x_i\right]$$

for any $x_1, \ldots, x_I$ in $[0, 1]$.

(iii) For any $p > 0$,

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\alpha_i^{(N)}(t))^p \stackrel{P}{\to}_N \mathbf{E}\left[\alpha(t)^p\right]. \tag{3.5}$$

Moreover, with initial conditions $q(0) = 0$ and $\alpha(0) = \alpha$, it holds that

$$q(t+1) \ = \ \left[q(t) - C + (1 - f(q(t)))\,\mathbf{E}\left[\alpha(t)\right]\right]^+ \tag{3.6}$$

and

$$\begin{aligned} \alpha(t+1) \ &=_{st} \ \alpha(t)^{1-\varepsilon}(1 - R(t+1))B(t+1) \\ &+ \ \gamma\alpha(t)R(t+1)B(t+1) \\ &+ \ \alpha(t)(1 - B(t+1)), \end{aligned} \tag{3.7}$$

where

$$B(t+1) = \mathbf{1}\left[U(t+1) \le \alpha(t)\right] \tag{3.8}$$

and

$$R(t+1) = \mathbf{1}\left[V(t+1) \le f(q(t))\right] \tag{3.9}$$

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

A proof of Theorem 1 is available in [59].

Theorem 1 suggests that during timeslot $[t, t+1)$ a bottleneck queue driven by a random dropping algorithm under a large number of TCP sources can be

characterized by a two-dimensional recursion for the limiting normalized queue length $q(t)$ and the limiting transmission rate $\alpha(t)$. The convergence result $\frac{Q^{(N)}(t)}{N} \xrightarrow{P}_N q(t)$ is a byproduct of the convergence

$$\frac{A^{(N)}(t)}{N} = \frac{\sum_{i=1}^{N} A_i^{(N)}(t)}{N} \xrightarrow{P}_N a(t) \tag{3.10}$$

with

$$a(t) = (1 - f(q(t)))\mathbf{E}\left[\alpha(t)\right].$$

This result, while similar to the classical Weak Law of Large Numbers, cannot be obtained by a straightforward application of the classical Law of Large Numbers. Indeed, the summands in (1.8) (under (AR2)) are *identically* distributed but *correlated* rvs whose common distribution *varies* with N. However, as the number of sources increases, the dependency between any pair of sources becomes weaker so that the aggregate behavior eventually becomes deterministic. Since the transmission rates are random and asymptotically independent, Theorem 1 provides some indications that the transmission rates among all flows are no longer synchronized when the number of flows is large. It also helps justify the use of micro-scale TCP flow models because any single flow has no impact to the global behavior of the system – The behavior of a flow can be decoupled from the system and the model only needs to take into account the reaction of a flow to the system but not vice versa.

One advantage of modeling a RED gateway over a Tail-Drop gateway is that in RED gateway, there is a fixed orderly structure on how the packets are being marked/dropped depending on the probability function. In a Tail-Drop gateway, it is more difficult to accurately model how incoming packets are dropped, for such events usually depend on the precise timing of packets arrivals and

departures. This absence of randomness in the Tail-Drop gateway can also give rise to a complex phenomenon such as chaotic behavior as suggested in [61].

3.3 Simulations of the Rate-based Model

In this section, we present the results from Monte-Carlo simulations of the ersatz model presented in Section 3.1 to demonstrate the asymptotic behavior derived in Theorem 1. We simulate the system for $N = 10, 100, 1000$ with $\varepsilon = 0.1$, $\gamma = 0.5$ and $C = 0.5$; the initial conditions are $Q^{(N)}(0) = 0$ and $\alpha_i^{(N)} = 0.5$ for all $i = 1, .., N$. The drop probability is calculated through the piecewise linear function $f : \mathbb{R}_+ \to [0, 1]$ given by

$$f(x) = \begin{cases} 0 & x < 1 \\ \frac{x-1}{4} & 1 \leq x < 5 \\ 1 & 5 \leq x. \end{cases} \tag{3.11}$$

The simulation results are shown in Figure 3.1 and 3.2. It is clear that the fluctuation of $Q^{(N)}(t)/N$ decreases as the number of sources increases, and the same is true for the average transmission rate. With a hundred or more flows, our analytical result seems to hold reasonably well. Moreover, this simulation result also suggests the existence of the steady-state, which happens quickly after only around a hundred iterations. Similar results obtained from a full-scale simulation of the protocol stacks in *NS* [45] are presented in the next chapter. These preliminary findings comfort our belief that the ersatz model captures some of the essential features of TCP and RED.

We are keenly aware that the ersatz proposed here fails to incorporate some key features of TCP and RED, most notably round trip delays and queue

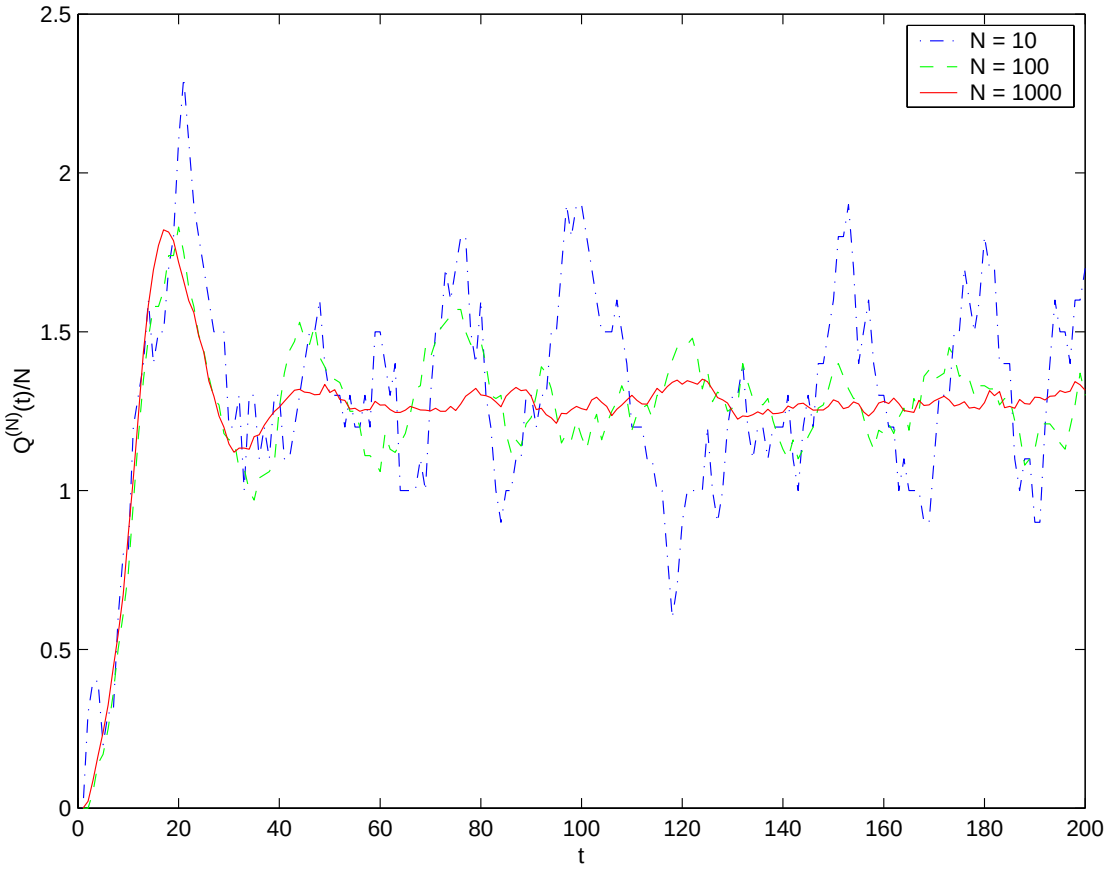

Figure 3.1: The normalized queue length of Simulation 1.

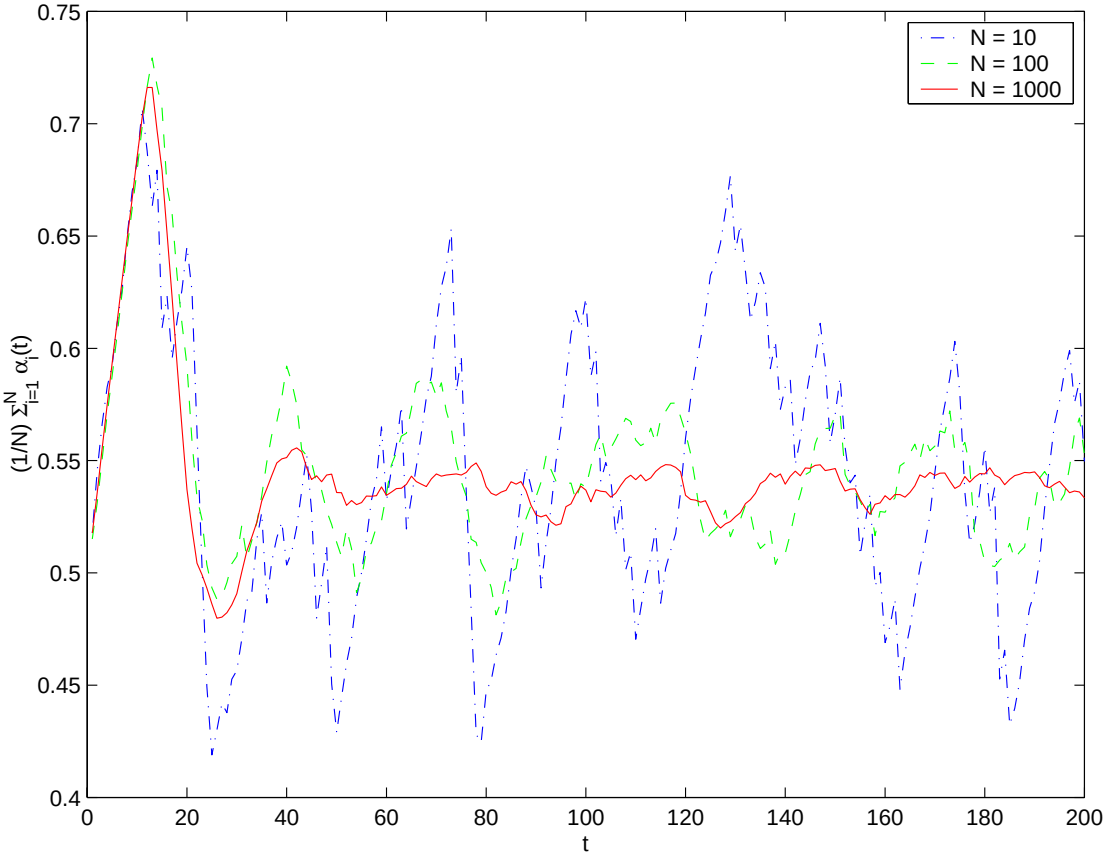

Figure 3.2: The average transmission rate per user of Simulation 1.

averaging. It is however noteworthy that despite such omissions, the asymptotics derived from this simplified model are qualitatively in line with the results obtained via the NS simulator (which emulates in great details the TCP stack). Thus, this qualitative convergence points to the possibility of creating simple and yet accurate models that capture key interactions between the TCP and RED mechanisms in the presence of a large number of TCP connections. Indeed, a more accurate model with a similar asymptotic behavior is presented in Chapter 4.

Chapter 4

A Window-based Model of Persistent TCP Population with Homogeneous Round-trip Delays

In this chapter, we embellish the rate-based model presented in Chapter 3 to fully incorporate the window mechanism of TCP congestion-control. As a result, each TCP flow can transmit more than one packet per timeslot unlike the ersatz model in Chapter 3. Another major difference is that TCP flows presented here use *marked* packets (*e.g.,* by Explicit Congestion Notification [15]) instead of *dropped* packets as a signaling mechanism for congestion-control. We take this approach for two reasons: (i) Since marking reduces the amount of unnecessary retransmissions and produces more desirable properties in network traffic, it is a signaling mechanism of choice for AQM traffic management; (ii) The actual window mechanism for TCP regarding dropped packets is complicated as the TCP source needs to receive triple-duplicated acknowledgments before triggering the congestion-control mechanism, hence the size of the congestion window and the order of the dropped packets in the timeslot can cause different behavior.

For this model, we establish a Weak Law of Large Numbers in Theorem 2, a Central Limit Theorem (CLT) in Theorem 3 and the steady-state analysis, all of

which reveal several interesting points:

(i) Theorem 2 shows that the dynamics of the queue at time t, still denoted $Q^{(N)}(t)$, can be approximated by $Nq(t)$ with $q(t)$ determined via a simple deterministic recursion, which is independent of the number of users. This approximation becomes more accurate as the number of users becomes large. The limiting model is therefore "scalable" as it does not suffer from state space explosion, nor does it require any ad-hoc assumptions.

(ii) Theorem 2 also shows that the dependency between each TCP flow becomes negligible under a large number of flows, *i.e.,* "RED breaks the global synchronization when the number of flows is large." More specifically, the window sizes of all users are asymptotically independent and random, hence the aggregate traffic becomes smooth unlike in Tail-Drop gateways where synchronization between flows causes the aggregate traffic to exhibit the sawtooth pattern.

(iii) The bottleneck capacity being NC, the average throughput of each flow is approximately C for large N, so that TCP traffic does not follow a commonly held belief associated with statistical multiplexing. Indeed, for N open-loop independent traffic sources operating at average rate C, statistical multiplexing suggests that the bandwidth required to serve the aggregate flow is less than NC. TCP flows, on the other hand, are correlated due to their coupling via the bottleneck router, and multiplexing TCP flows is not as effective.

(iv) The queue length in steady-state can be easily calculated, while the steady-state distribution of the window size and the average window size can be evaluated from well-known TCP models.

(v) For a more accurate description of the queue dynamics, a Central Limit Theorem-type analysis (summarized in Theorem 3) yields the existence of a rv $L_0(t)$ such that the approximation (1.7) holds.

(vi) This CLT analysis also reveals that the magnitude of queue fluctuations is proportional to the derivative of the marking probability function. This advocates the use of a smooth marking probability function in line with the RED "gentle" option recently suggested as opposed to the original recommendation in [16, 18]. This finding coincides with the observed oscillatory behavior of RED when the average packet drop rate exceeds max_p, in the absence of RED's "gentle" modification [14].

The chapter is organized as follows. In Section 4.1, the model is described in details, and a first set of asymptotic results are presented in Section 4.2. Section 4.3 focuses on the calculations of the limiting normalized queue size and the average window size in steady state. Section 4.4 contains a Central Limit Theorem complement to the asymptotic results of Section 4.2. Simulation results confirming the theoretical results are shown in Section 4.5. Applications to network dimensioning and to marking probability function design are demonstrated in Section 4.6.

4.1 Model Description

Overview of the dynamics

The TCP congestion-control algorithm dynamically adjusts the size of the congestion window (the amount of unacknowledged packets in the network per

round-trip) by the following mechanism (assuming ECN/RED is utilized at the bottleneck node): If all the packets transmitted in a round-trip are not marked, then the size of the congestion window is increased by one packet for the next round-trip. On the other hand, if at least one packet is marked in the round-trip, the congestion window is halved. The probability that the router will mark an incoming packet in the buffer depends on the average queue length at the time of its arrival. The average queue length is calculated by an exponential weighted average filter with large time-constant to prevent RED from reacting too fast. As a result, consecutive incoming packets into RED are marked with almost identical probability. With this in mind, we now construct a model whose dynamics are similar in spirit to the dynamics of TCP + ECN/RED.

The discrete-time model

Similarly to the model in Chapter 3, time is assumed discrete and slotted in contiguous timeslots of duration equal to the round-trip delay of TCP connections. We consider N traffic sources, all transmitting through a bottleneck RED gateway with ECN enabled in both TCP and RED. The capacity of this bottleneck link is NC packets/slot for some positive constant C. The RED buffer is modeled as an infinite queue, so that no packet losses occur due to buffer overflow, and congestion-control is achieved solely through the random marking algorithm in the RED gateway.

Dynamics

Fix $N = 1, 2, \ldots$, and suppose that each of the N sources (*i.e.,* TCP connections) has an infinite amount of data to transmit and that in each timeslot it transmits

as much as allowed by its congestion window in that timeslot. So, for $i = 1, \ldots, N$, let $W_i^{(N)}(t)$ be an integer-valued rv that encodes the number of packets generated by source i (and hence its congestion window) at the beginning of the timeslot $[t, t+1)$. We assume the integer $W_i^{(N)}(t)$ to be in the range $\{1, \ldots, W_{\max}\}$ for some finite integer $W_{\max}$. Throughout we assume $W_{\max} \geq 2$ to avoid boundary cases of limited interest.

Upon arrival at the RED gateway, each packet from source i may be marked according to a random marking algorithm (to be specified shortly). We represent this possibility by the $\{0,1\}$-valued rv $M_{i,j}^{(N)}(t+1)$ $(j = 1, ..., W_i^{(N)}(t))$ 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. Given that N sources are active, the total number of packets which are accepted into the RED buffer at the beginning of timeslot $[t, t+1)$ is given by (1.8), where in this model

$$A_i^{(N)}(t) := W_i^{(N)}(t). \tag{4.1}$$

Again, the dynamics of the queue size is given as (1.9), which can be rewritten as

$$Q^{(N)}(t+1) = \left[Q^{(N)}(t) - NC + \sum_{i=1}^{N} W_i^{(N)}(t) \right]^+. \tag{4.2}$$

Statistical assumptions

In order to fully specify the model, we need to specify the *joint* statistics of the rvs $\{M_{i,j}^{(N)}(t+1), A_i^{(N)}(t),\ i = 1, \ldots, N;\ j = 1, 2, \ldots;\ t = 0, 1, \ldots\}$. To do so we introduce the collection of i.i.d. $[0,1]$-*uniform* rvs $\{V_i(t+1),\ V_{i,j}(t+1),\ i, j = 1, \ldots;\ t = 0, 1, \ldots\}$ which are assumed independent

of the rvs $A_1^{(N)}(0), \ldots, A_N^{(N)}(0)$ and $Q^{(N)}(0)$. We also introduce a mapping $f^{(N)} : \mathbb{R}_+ \to [0,1]$ which acts as the *marking probability* function of the RED gateway.

The process by which packets are marked is described first: For each $i = 1, \ldots, N$, we define the marking rvs

$$M_{i,j}^{(N)}(t+1) \;=\; \mathbf{1}\left[V_{i,j}(t+1) > f^{(N)}(Q^{(N)}(t))\right], \quad j = 1, 2, \ldots \tag{4.3}$$

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 the timeslot $[t, t+1)$. Thus, in a round-trip, each packet coming into the router is marked with identical (conditional) probability which depends only on the queue length at the beginning of the timeslot. This model approximates the case where the memory of the queue averaging mechanism is long, which is the case for the recommended parameter settings of RED [16].

Next we introduce the rvs

$$M_i^{(N)}(t+1) = \prod_{j=1}^{A_i^{(N)}(t)} M_{i,j}^{(N)}(t+1) \tag{4.4}$$

so that $M_i^{(N)}(t+1) = 1$ (resp. $M_i^{(N)}(t+1) = 0$) corresponds to the event that no packet (resp. at least one packet) from source i has been marked in timeslot $[t, t+1)$. Recall that in this model, $A_i^{(N)}(t) = W_i^{(N)}(t)$, *i.e.,* each connection injects as many packets into the network as its congestion window will allow. The evolution of the window mechanism for source i can now be described through the recursion

$$\begin{aligned} W_i^{(N)}(t+1) \;=\; & \min\left(W_i^{(N)}(t)+1, W_{\max}\right) M_i^{(N)}(t+1) \\ & + \min\left(\lceil \frac{W_i^{(N)}(t)}{2} \rceil, W_{\max}\right)(1 - M_i^{(N)}(t+1)). \end{aligned} \tag{4.5}$$

This equation emulates the interaction between TCP and RED as follows: If no packet from source i is marked in the timeslot $[t, t+1)$, then the congestion window size in the next timeslot is increased by one packet. On the other hand, if one or more packets are marked in the timeslot $[t, t+1)$, then the congestion window in the next timeslot is reduced by half. The size of the congestion window is limited by the maximum window size $W_{\max}$ [1].

4.2 The Asymptotics

The first result of this model consists in the asymptotics for the normalized buffer content as the number of TCP flows becomes large. This result, contained in Theorem 2, is discussed under the following assumptions (AW1)-(AW2):

(AW1) There exists a continuous function $f : \mathbb{R}_+ \to [0, 1]$ such that for each $N = 1, 2, \ldots$, Equation (1.1) holds.

(AW2) For each $N = 1, 2, \ldots$, the dynamics (4.2) and (4.5) start with the conditions

$$Q^{(N)}(0) = 0 \quad \text{and} \quad W_i^{(N)}(0) = W, \quad i = 1, \ldots, N$$

for some integer W in the range $\{1, \ldots, W_{\max}\}$.

Note that these assumptions are similar in nature to Assumption (AR1)-(AR2) in Chapter 3.

[1] Note also that if $W_i^{(N)}(0)$ lies in the range $\{1, \ldots, W_{\max}\}$ for each $i = 1, \ldots, N$, then so does $W_i^{(N)}(t)$ for each $t = 0, 1, \ldots$ and the minimum with $W_{\max}$ in the second term of (4.5) can be omitted.

Theorem 2. *Assume (AW1)-(AW2) to hold. Then, for each $t = 0, 1, \ldots$, there exist a (non-random) constant $q(t)$ and an $\{1, \ldots, W_{\max}\}$-valued rv $W(t)$ such that the following holds:*

(i) The convergence results

$$\frac{Q^{(N)}(t)}{N} \stackrel{P}{\to}_N q(t) \quad \text{and} \quad W_1^{(N)}(t) \Longrightarrow_N W(t) \tag{4.6}$$

take place;

(ii) For any function $g : \mathbb{N} \to \mathbb{R}$,

$$\tfrac{1}{N} \textstyle\sum_{i=1}^{N} g(W_i^{(N)}(t)) \stackrel{P}{\to}_N \mathbf{E}\left[g(W(t))\right]. \tag{4.7}$$

(iii) For any integer $I = 1, 2, \ldots$, the rvs $\{W_i^{(N)}(t),\ i = 1, \ldots, I\}$ become asymptotically independent as N becomes large, with

$$\lim_{N\to\infty} \mathbf{P}\left[W_i^{(N)}(t) = k_i,\ i = 1, \ldots, I\right] \ = \ \prod_{i=1}^{I} \mathbf{P}\left[W(t) = k_i\right] \tag{4.8}$$

for any $k_1, \ldots, k_I$ in $\mathbb{N}$

Moreover, with initial conditions $q(0) = 0$ and $W(0) = W$, it holds that

$$q(t+1) = \left(q(t) - C + \mathbf{E}\left[W(t)\right]\right)^+ \tag{4.9}$$

and

$$\begin{aligned} W(t+1) \ =_{st} \ & \min\left(W(t)+1, W_{\max}\right) M(t+1) \\ & + \min\left(\lceil \frac{W(t)}{2} \rceil, W_{\max}\right)\left(1 - M(t+1)\right), \end{aligned} \tag{4.10}$$

where

$$M(t+1) = \mathbf{1}\left[V(t+1) \le \left(1 - f(q(t))\right)^{W(t)}\right] \tag{4.11}$$

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

A proof of Theorem 2 is available in Chapter 6. As should be clear from the discussion given there, Theorem 2 readily flows from a Weak Law of Large Numbers (4.7) for the triangular array

$$\{W_i^{(N)}(t),\ i = 1, \ldots, N; N = 1, 2, \ldots\}. \tag{4.12}$$

The numerical calculations for the limiting model are very simple. The number of states required for the calculation for each time step is only $W_{\max} + 1$ regardless of N. For each $t = 0, 1, \ldots$, we can determine $q(t)$ through the following steps:

(i) For $t = 0$, start with some given values $q(0) = 0$ and $W(0) = W$, *i.e.*, $\mathbf{P}[W(0) = j] = \delta(j, W)$, $j = 1, \ldots, W_{\max}$, and use $\mathbf{E}[W(0)] = W$ to calculate $q(1)$ via (4.9);

(ii) Given $q(t)$ and $\mathbf{P}[W(t) = j]$, $j = 1, \ldots, W_{\max}$, for some $t = 0, 1, \ldots$, use (4.10) and (4.11) with $q(t)$ to calculate the transition probabilities and $\mathbf{P}[W(t+1) = j], j = 1, \ldots, W_{\max}$. Then calculate $\mathbf{E}[W(t+1)]$;

(iii) Use $\mathbf{E}[W(t+1)]$ in (ii) to update $q(t+2)$ from (4.9);

(iv) Increase t by one and repeat Step (ii)-(iv).

Examples of the numerical calculation of the limiting model will be presented in Section 4.5.

4.3 Steady-state Regime

We now turn our attention to the steady state regime of the limiting two-dimensional recursion (4.9)-(4.11), more specifically to the calculation of the

limiting queue and average window size in statistical equilibrium, *i.e.,* large t asymptotics. We show that there is a close relationship between this limiting behavior in steady state and the now standard TCP throughput model with stationary loss developed in [47, 48]:

Throughout, we assume the following assumptions (AW3)-(AW4), where

(AW3) The marking function $f : \mathbb{R} \to [0, 1]$ is increasing with

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

(AW4) The sequence $\{(q(t), W(t)),\ t = 0, 1, \ldots\}$ admits a steady state in the sense that

$$(q(t), W(t)) \Longrightarrow_t (q^\star, W^\star)$$

for some rvs $(q^\star, W^\star)$ where $q^\star$ is non-random and $W^\star$ is an $\{1, \ldots, W_{\max}\}$-valued rv.

Although the two-dimensional sequence $\{(q(t), W(t)),\ t = 0, 1, \ldots\}$ is a time-homogeneous Markov chain with values in $\mathbb{R}_+ \times \{1, \ldots, W_{\max}\}$, we shall not address here the existence of the limit posted in (AW4) as complications arise due the fact that the first component is degenerate (*i.e.,* deterministic). Control-theoretic analyses (as surveyed in Section 2.2) suggest that the existence of the limit in the sense of (AW4) depends on the parameter settings of RED. If the parameters are set appropriately, then the queue will stabilize to its fixed-point solution. Otherwise, the steady-state queue can fluctuate around its fixed-point solution (due to border collisions) or can even exhibit chaotic behavior. In any case, only the steady-state in the sense of (AW4) is desirable. Discussions on the parameter settings in RED to ensure (AW4), *i.e.,* stability, can be found in Section 2.2.

However, that under the condition $W_{\max} \leq C$, it is a simple matter to check that (AW4) holds with $q^\star = 0$ and $W^\star = W_{\max}$ as would be expected; related details are given in Section 4.3.2 (Case 1). Thus, only the case $C < W_{\max}$ needs to be considered.

4.3.1 TCP throughput with fixed loss probability

In order to discuss the limit $(q^\star, W^\star)$ postulated in (AW4), we first need to review the results from the TCP throughput model with stationary loss [47, 48]:

Fix γ in $[0, 1]$, where $\gamma = 1 - p$ with p interpreted as the stationary dropping/marking probability. Consider the recursion

$$\begin{aligned} W_0^\gamma &= W_0; \\ W_{n+1}^\gamma &= \min\left(W_n^\gamma + 1, W_{\max}\right) M_{n+1}^\gamma + \min\left(\lceil \frac{W_n^\gamma}{2} \rceil, W_{\max}\right)(1 - M_{n+1}^\gamma) \end{aligned}$$

for all $n = 0, 1, \ldots$ and some $\{1, \ldots, W_{\max}\}$-valued rv W_0, with

$$M_{n+1}^\gamma = \mathbf{1}\left[V_{n+1} \leq \gamma^{W_n^\gamma}\right], \quad n = 0, 1, \ldots \tag{4.13}$$

where the rvs $\{V_{n+1},\ n = 0, 1, \ldots\}$ are i.i.d. $[0, 1]$-uniform rvs which are independent of W_0.

The rvs $\{W_n^\gamma,\ n = 0, 1, \ldots\}$ form a time-homogeneous Markov chain on the finite set $\{1, \ldots, W_{\max}\}$. For γ in the open interval $(0, 1)$, this chain is irreducible, positive recurrent and aperiodic, thus ergodic. Consequently,

$$W_n^\gamma \Longrightarrow_n W^\gamma \tag{4.14}$$

for some $\{1, \ldots, W_{\max}\}$-valued rv W^γ. This rv W^γ satisfies the distributional equation

$$W^\gamma =_{st} \min\left(W^\gamma + 1, , W_{\max}\right) M^\gamma + \min\left(\lceil \frac{W^\gamma}{2} \rceil, W_{\max}\right)(1 - M^\gamma) \tag{4.15}$$

where

$$M^\gamma = \mathbf{1}\left[V \leq \gamma^{W^\gamma}\right] \tag{4.16}$$

for some $[0,1]$-uniform rv V which is independent of the rv W^γ. In fact, the ergodicity of the Markov chain guarantees that the equation (4.15)-(4.16) has a solution and that this solution is unique.

For the sake of completeness, we also consider the boundary cases: For $\gamma = 0$ (resp. $\gamma = 1$), it is easy to see that (4.14) also takes place with $W^\gamma = 1$ (resp. $W^\gamma = W_{\max}$). The converse is also true as we now demonstrate: If $W^\gamma = 1$ under (4.14), then (4.15)-(4.16) read

$$1 =_{st} \min\left(1 + M^\gamma, W_{\max}\right)$$

with $M^\gamma = \mathbf{1}\left[V \leq \gamma\right]$, so that necessarily $M^\gamma = 0$ under (AW4), whence $\gamma = 0$.

On the other hand, if $W^\gamma = W_{\max}$ under (4.14), then (4.15)-(4.16) now reduce to

$$W_{\max} =_{st} W_{\max} M^\gamma + \lceil \frac{W_{\max}}{2} \rceil (1 - M^\gamma)$$

with

$$M^\gamma = \mathbf{1}\left[V \leq \gamma^{W_{\max}}\right].$$

Consequently,

$$\lfloor \frac{W_{\max}}{2} \rfloor =_{st} \lfloor \frac{W_{\max}}{2} \rfloor M^\gamma.$$

so that $M^\gamma = 1$, whence $\gamma = 1$.

Although we can numerically compute the steady-state distribution determined by (4.15)-(4.16) (which is a special case of [47]), we take notice that this model is actually that of a single TCP connection with a constant loss probability $1 - \gamma$. This is a well-studied problem (*e.g.,* [40, 48]) with known

results. If we replace $\lfloor \frac{W_n^\gamma}{2} \rfloor$ in (4.15) with $\frac{W_n^\gamma}{2}$, then we can invoke Eqn. (33) in [48] to get the approximation

$$\mathbf{E}\left[W^\gamma\right] \simeq \min\left(W_{\max}, \sqrt{\frac{3}{2(1-\gamma)}}\right). \tag{4.17}$$

4.3.2 Steady-state regime for the model in Section 4.1

Under Assumption (AW4), it is a simple matter to see that

$$(q(t), W(t), M(t+1)) \Longrightarrow_t (q^\star, W^\star, M^\star)$$

with

$$M^\star = \mathbf{1}\left[V \leq (1 - f(q^\star))^{W^\star}\right] \tag{4.18}$$

where the $[0,1]$-uniform rv V is independent of the pair $(q^\star, W^\star)$.

Upon letting t go to infinity in (4.9) and (4.10), we obtain the relations

$$q^\star = (q^\star - C + \mathbf{E}\left[W^\star\right])^+ \tag{4.19}$$

and

$$W^\star \quad =_{st} \quad \min\left(W^\star + 1, W_{\max}\right) M^\star + \min\left(\lceil \frac{W^\star}{2} \rceil, W_{\max}\right)(1 - M^\star). \tag{4.20}$$

With $q^\star$ given, the solution to (4.20) with (4.18) exists and is unique; it is in fact given by

$$W^\star = W^\gamma \quad \text{with} \quad \gamma = 1 - f(q^\star)$$

where the rv W^γ is defined through (4.14). Several cases are possible when considering (4.19) and (4.20).

Case 1 – $f(q^\star) = 0$: Then, $q^\star$ is finite under (AW3), $M^\star = 1$ and (4.20) reduces to

$$W^\star =_{st} \min\left(W^\star + 1, W_{\max}\right)$$

with unique solution $W^\star = W_{\max}$ (in agreement with the discussion in Section 4.3.1). In that case, (4.19) gives

$$q^\star = (q^\star - C + W_{\max})^+ . \tag{4.21}$$

Therefore, either $q^\star = 0$ in which case $W_{\max} \leq C$ as should be expected, or $q^\star > 0$ (still with $f(q^\star) = 0$), in which case $C = W_{\max}$. However, under the condition $W_{\max} \leq C$, it is clear that if $Q^{(N)}(0) = Q > 0$ for all $N = 1, 2, \ldots$, then $Q^{(N)}(t) \leq Q$ and the conclusion $q(t) = 0$ holds for all $t = 0, 1, \ldots$, so that $q^\star = 0$. In other words, if $q^\star$ in (AW4) is such that $f(q^\star) = 0$, then necessarily $q^\star = 0$. □

Case 2 – $f(q^\star) = 1$: Then $q^\star > 0$, $M^\star = 0$ and (4.20) now reduces to

$$W^\star =_{st} \min\left(\lceil \frac{W^\star}{2} \rceil, W_{\max} \right)$$

with only solution $W^\star = 1$ (also in agreement with the discussion in Section 4.3.1). □

Case 3 – $0 < f(q^\star) < 1$: Then, $0 < q^\star < \infty$ and from (4.19) it is necessarily the case that $\mathbf{E}[W^\star] = C$. Thus, the existence of a steady state requires at the very least that the equation

$$\mathbf{E}[W^\gamma] = C, \quad \gamma \in [0, 1] \tag{4.22}$$

has a unique solution, say $\gamma^\star$, in which case we must have

$$\gamma^\star = 1 - f(q^\star).$$

By known results on finite state Markov chains, the mapping $\gamma \to \mathbf{E}[W^\gamma]$ is continuous on $[0, 1]$ [39] with

$$\mathbf{E}[W^\gamma]_{\gamma=0} = 1 \quad \text{and} \quad \mathbf{E}[W^\gamma]_{\gamma=1} = W_{\max}.$$

By continuity, $\{\mathbf{E}[W^\gamma],\ \gamma \in [0,1]\}$ must contain the interval $[1, W_{\max}]$. On the other hand, it is always the case that

$$1 \leq \mathbf{E}[W^\gamma] \leq W_{\max}, \quad \gamma \in [0,1],$$

and we conclude that $\{\mathbf{E}[W^\gamma],\ \gamma \in [0,1]\} = [1, W_{\max}]$. Therefore, there exists at least one solution to (4.22) *provided* $1 \leq C \leq W_{\max}$. The uniqueness of the solution would be ensured by the strict monotonicity of the mapping $\gamma \to \mathbf{E}[W^\gamma]$. Although we have not been able to establish the monotonicity, it is suggested by the approximation (4.17) and also from the simulation result illustrated in Figure 4.1. □

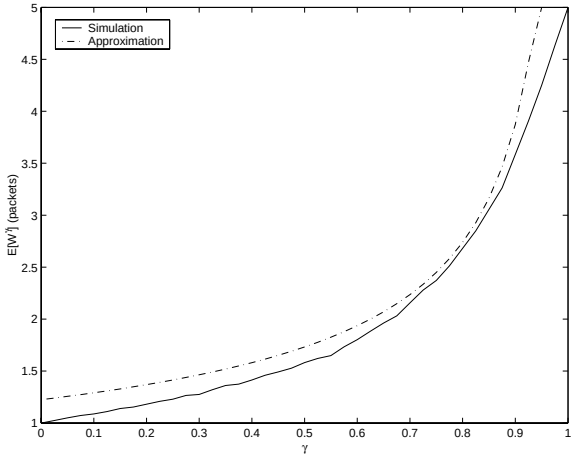

Figure 4.1: The mapping $\gamma \to \mathbf{E}[W^\gamma]$ when $W_{\max} = 5$ vs. the approximation (4.17).

4.3.3 Discussion

The preceding analysis demonstrates that the resulting steady-state queue and congestion window can be determined by the steady-state marking probability

$f(q^\star)$. We now discuss different cases that arise depending on the value $f(q^\star)$.

The trivial case $f(q^\star) = 0$ is possible if and only if the maximum window size $W_{\max} \leq C$, or equivalently, no congestion exists and congestion-control is not necessary. On the other hand, when $f(q^\star) = 1$ the window size will be reduced to its minimum value (which is one packet in this model). Even then, the congestion is so severe that the limiting queue will grow unboundedly large. This case is possible if and only if $C < 1$, and this is clearly not a desirable operating regime for the system.

In the non-trivial case where $0 < f(q^\star) < 1$, the average throughput will be equal to the capacity per flow C. Furthermore, $\mathbf{E}\left[W^{1-f(q^\star)}\right] = C$ where $W^{1-f(q^\star)}$ is the steady-state rv of the Markov chain in (4.15) when $\gamma = 1 - f(q^\star)$. In other words, the throughput of a flow in steady state can be evaluated as a TCP flow with fixed-loss probability $f(q^\star)$. Conversely, assuming C is known, the (assumed) monotonicity of the mapping $\gamma \to \mathbf{E}\left[W^\gamma\right]$ along with (4.22) can be used to evaluate the steady-state marking probability $f(q^\star)$ and the steady-state queue length $q^\star$ (assuming f is invertible). This application to network dimensioning is discussed in more details in Section 4.6.

Finally, we note that the steady-state analysis implies that the behavior of any single TCP flow is *decoupled* from the dynamics of the system when the number of flows is large, *i.e.,* the behavior of a single flow no longer affects the dynamics of the system. This enables the performance analysis of the system to be done at two levels: (i) From a user's perspective, the network only generates a fixed marking/signaling rate to a TCP flow and the performance of a TCP flow can be evaluated from the TCP throughput formula such as in [48]; and (ii) From the network operator's perspective, the aggregate traffic and its interaction with

AQM can be easily predicted without having to consider any particular flows.

4.4 A Central Limit Theorem

In this section, we present a Central Limit Theorem (CLT) which complements the limiting results obtained earlier. The discussion is carried out in the setup of Section 4.2, but with Assumption (AW1) strengthened to read as Assumption (AW1b), where

(AW1b) Assumption (AW1) holds with continuously differentiable mapping $f : \mathbb{R}_+ \to [0,1]$, *i.e.,* the derivative $f' : \mathbb{R}_+ \to \mathbb{R}$ exists and is continuous.

For each $t = 0, 1, \ldots$, let $q(t)$ and $W(t)$ be as in Theorem 2, and set

$$L_0^{(N)}(t) := \frac{Q^{(N)}(t)}{N} - q(t) \tag{4.23}$$

and

$$\bar{L}^{(N)}(t) = \frac{1}{N}\left(\sum_{i=1}^{N} W_i^{(N)}(t) - \mathbf{E}\left[W(t)\right]\right). \tag{4.24}$$

Theorem 3. *Assume (AW1b)-(AW2) to hold. Then, for each $t = 0, 1, \ldots$, there exists an $\mathbb{R}^2$-valued rv $\mathbf{L}(t) = (L_0(t), \bar{L}(t))$ such that the convergence*

$$\sqrt{N}\left(L_0^{(N)}(t), \bar{L}^{(N)}(t)\right) \Longrightarrow_N \mathbf{L}(t) \tag{4.25}$$

holds. Moreover, the distributional recurrence

$$L_0(t+1) =_{st} \begin{cases} 0 & K(t) > 0 \\ L_0(t) + \bar{L}(t) & K(t) < 0 \\ \left(L_0(t) + \bar{L}(t)\right)^+ & K(t) = 0 \end{cases} \tag{4.26}$$

holds where we have set

$$K(t) = C - q(t) - \mathbf{E}\left[W(t)\right]. \tag{4.27}$$

The convergence (4.25) suggests the approximation (1.7) and

$$\sum_{i=1}^{N} W_i^{(N)}(t) \simeq N\mathbf{E}\left[W(t)\right] + \sqrt{N}\bar{L}(t) \tag{4.28}$$

for large N.

We can interpret $K(t)$ as the residual capacity per user in the limit in the timeslot $[t, t+1)$. If there exists extra capacity for the average user rate to increase ($K(t) > 0$), then there is no fluctuation in the limiting queue. On the other hand, when there is congestion ($K(t) < 0$), the (non-trivial) limiting distribution can be found recursively. Some technical difficulties arise in the special case $K(t) = 0$.

The proof of Theorem 3 is given in Chapter 7, and relies on showing that some key convergence statements propagate over time. If we specialize (7.6), one of the by-product of this analysis, to the mapping $g : \mathbb{N} \to \mathbb{R}$ given by $g(w) = w$, we find that $\bar{L}(t+1)$ is of the form

$$\bar{L}(t+1) =_{st} \Lambda(t) + f'(q(t))R(t)L_0(t) + H(t+1) \tag{4.29}$$

where $R(t)$ is a constant, $H(t+1)$ is a zero-mean Gaussian rv independent of the pair of rvs $(L_0(t), \Lambda(t))$ and the statistics of rv $\Lambda(t)$ are determined by $q(0), q(1), \ldots, q(t-1)$.

From (4.29), we observe that the magnitude of the queue fluctuation is proportional to the derivative of f around the limiting (normalized) queue size

$q(t)$. Since the original drop probability function of RED is discontinuous around *max_thresh*, our analysis is very much in line with the reported oscillatory queue behavior when the average queue exceeds *max_thresh* and with the finding that the addition of the "gentle" option can improve the queue behavior [14].

To get an intuition on why the derivative of the marking function plays such an important role, note that $f(\frac{Q^{(N)}(t)}{N})$ is the only feedback information from the RED gateway to the TCP sources. However, this feedback information also fluctuates around its limiting mean $f(q(t))$. It is easy to imagine that the uncertainty in the feedback information will also lead to fluctuations in the limiting queue as is suggested by the following result which is a well-known result called the *Delta method* [60, Thm. 3.1, p.26]).

Proposition 1. *Let $f : \mathbb{R}_+ \to [0,1]$ be a differentiable mapping with derivative $f' : \mathbb{R}_+ \to \mathbb{R}$ continuous at $x = q(t)$. If for some $t = 0, 1, \ldots$, the convergence*

$$\sqrt{N}\left(\frac{Q^{(N)}(t)}{N} - q(t)\right) \Longrightarrow_N L_0(t) \tag{4.30}$$

takes place with some rv $L_0(t)$, then

$$\sqrt{N}\left(f\left(\frac{Q^{(N)}(t)}{N}\right) - f(q(t))\right) \Longrightarrow_N f'(q(t))L_0(t). \tag{4.31}$$

4.5 Simulations

In this section, we present results from (Monte-Carlo) simulations of the model presented in Section 4.1 *and* from NS simulations [45] to illustrate the behaviors suggested by both Theorems 2 and 3. For the NS simulations, we use the system shown in Figure 4.2. Each server establishes a TCP Reno connection to a

corresponding client, thereby competing for the capacity in the ECN/RED gateway. Each TCP has a fixed packet size of 1500 bytes and a maximum window size of 200 packets. The marking probability function in the ECN/RED gateway is specified as in (AW1) with $f : \mathbb{R}_+ \to [0, 1]$ taken to be

$$f(x) = \min\left(0.01(x-1)^+, 1\right), \quad x \geq 0.$$

We choose this simple marking probability function with only one major slope in order to later demonstrate the relationship between the magnitude of the queue fluctuations and the slope of f as mentioned in Section 4.4.

The "time constant" parameter w_q for the Exponential Weighted Moving Average is set to 0.002, similar to the recommended value in [16]. Every round equals the round-trip propagation delay of 200 milliseconds. At the beginning of each round, we collect the instantaneous queue length in the ECN/RED buffer for a total duration of 200 seconds. Figure 4.3 shows the queue length normalized by the number of connection (N) as a function of time. We note a behavior similar to that discussed in Theorem 2 as fluctuations in the normalized queue length decrease with an increasing number of connections. Moreover, Figure 4.3 also suggests the existence of a steady-state for the limiting model, with a steady-state normalized queue length being constant at approximately 4.85 packets/user, corresponding to the steady-state marking probability of $0.0385 = f(4.85)$.

To simulate the model described in Section 4.1, we use the same parameter setup as in the NS simulation, *i.e.,* $W_{\max} = 200$, simulation time of 1000 timeslots and the same marking function. The capacity per user (C) of the bottleneck router can be calculated from (4.17) and (4.22) when the marking probability p is 0.0385 (obtained from the NS simulation, so that $\gamma = 1 - p = .9615$). A simple

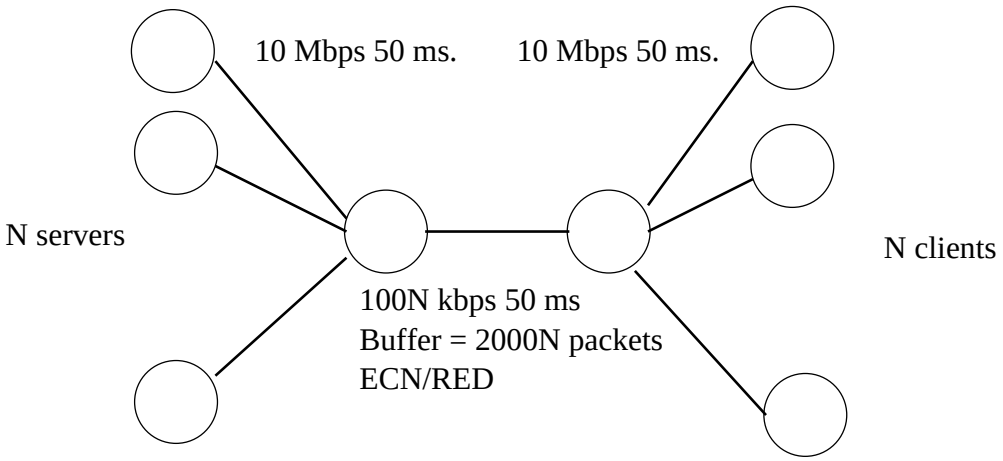

Figure 4.2: The topology setup in NS simulation.

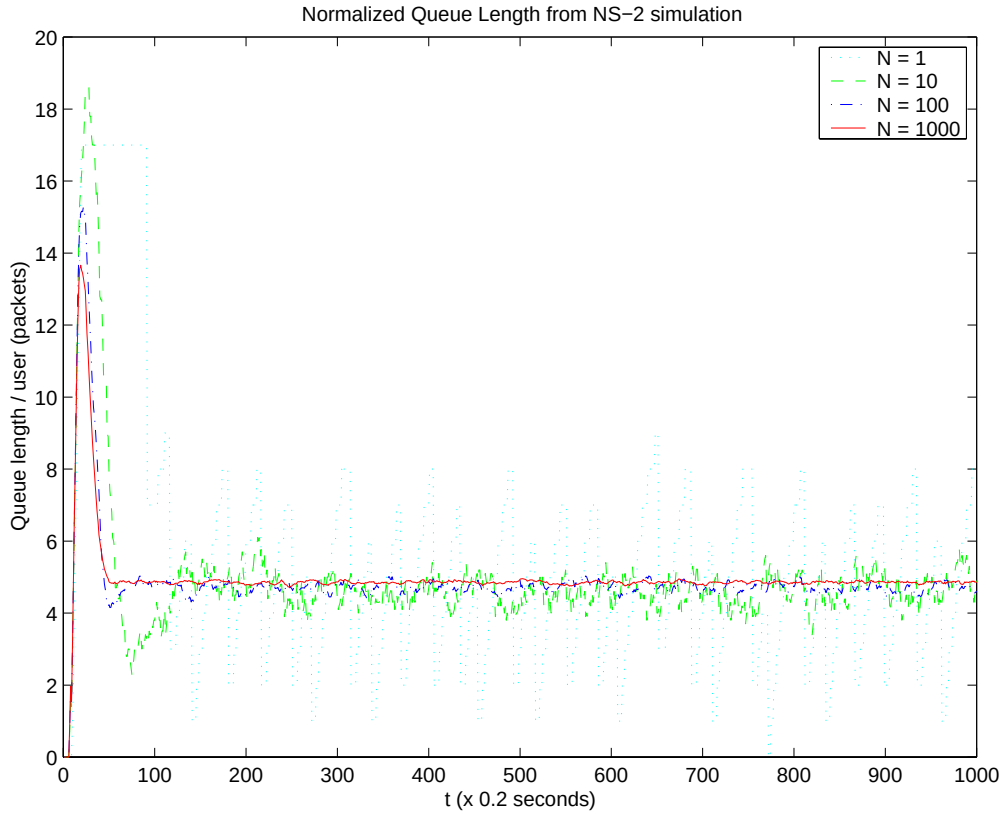

Figure 4.3: The normalized queue length of the ECN/RED gateway in NS simulation.

calculation yields $C = 6.24$ packets/timeslot. The simulation results are shown in Figure 4.4. A quick comparison to Figure 4.3 indicates a qualitative similarity in that the fluctuations decrease as the number of users increases. Further inspection reveals that the average normalized queue length is around 4.93 packets/user, very close to 4.85 packets/user produced by the NS simulation. Therefore, the limiting stochastic model appears to capture the essential behavior of queue dynamics in ECN/RED gateways, although the model exhibits somewhat greater fluctuations than in the NS simulation. This is due to the fact that all flows in the model are synchronized at the beginning of each timeslot, *i.e.,* they adapt at the same time while in the NS simulator, the flows react asynchronously to the marks from the RED gateway.

To gauge the rate of convergence, we assume that the queue is in steady state after the first 100 samples and that the magnitude of the queue fluctuation around its steady state mean is Gaussian and ergodic. Therefore, the steady-state standard deviation of the queue can be reasonably approximated from the sample standard deviation of the queue at timeslot 101 and beyond. The comparison between the sample standard deviation from the model and NS simulations is displayed in Figure 4.5. They clearly follow a similar trend. We also expect from Theorem 3 that the standard deviation will decrease as $1/\sqrt{N}$ for large N. Let S_N denote the sample standard deviation of the normalized queue when the number of users is N. We can see from Figure 4.5 that $S_1/\sqrt{N}$ provides a good approximation of the standard deviation S_N for large N.

To demonstrate the relationship between the slope of the marking probability function and the magnitude of the queue fluctuation (as mentioned in Section 4.4), we use the same network setup as before but with a slope increased ten-fold,

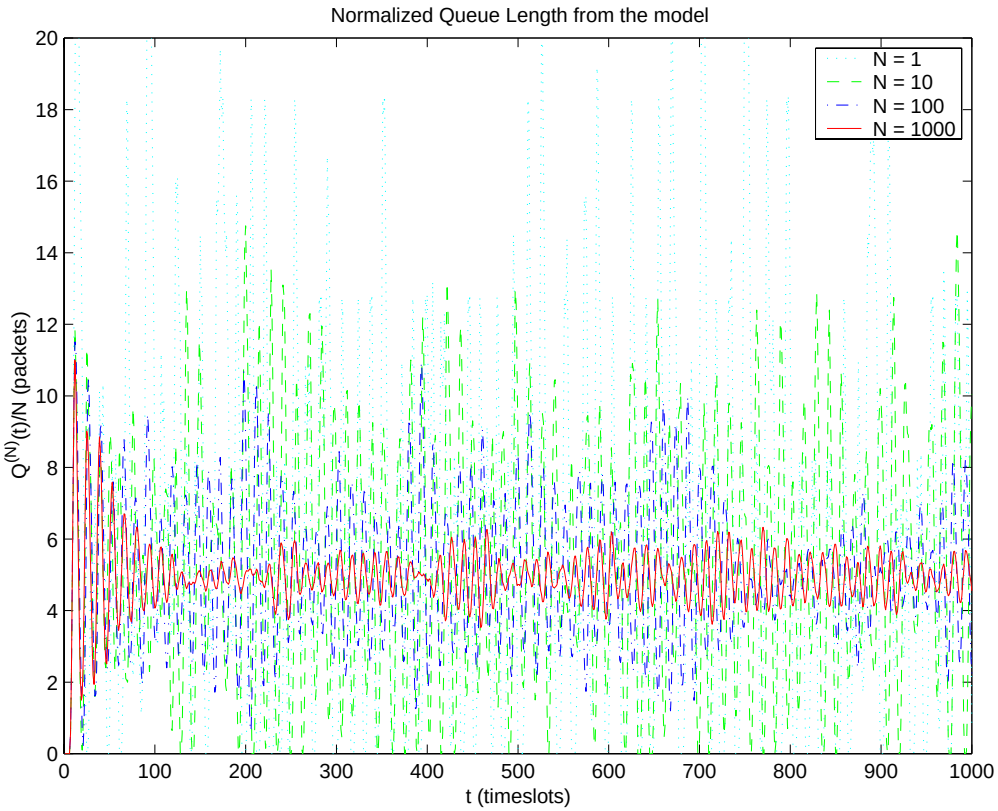

Figure 4.4: The normalized queue length of the model.

i.e.,

$$f(x) = \min\left(0.1(x-1)^+, 1\right), \quad x \geq 0. \tag{4.32}$$

Figure 4.6 and 4.7 show the simulation results in the case of the steeper function (4.32). In both the Monte-Carlo and NS simulations it is clear that the magnitude of queue fluctuation is much larger than in the original simulation. As the number of flows increases, the convergence becomes much slower than in the original setup. In the control-theoretic view of [22], this phenomenon can be interpreted as the control system becoming oscillatory with too large a feedback gain.

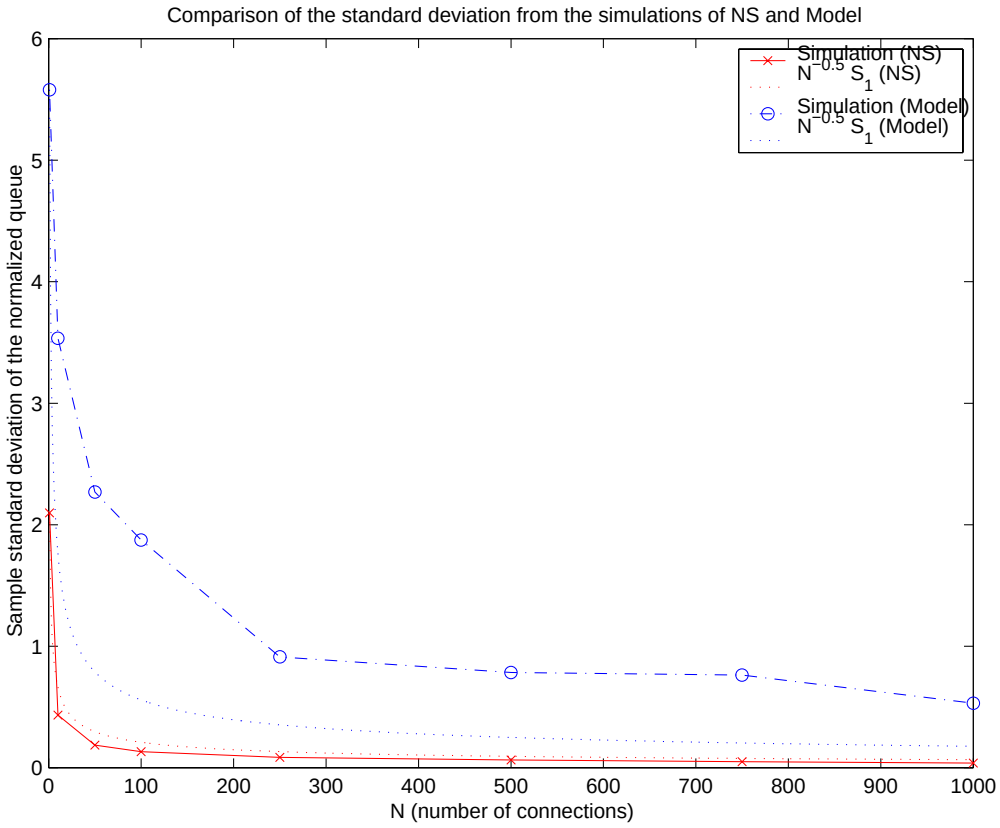

Figure 4.5: Sample standard deviation of the normalized queue.

4.6 Simple Network Dimensioning

We briefly discuss how to apply the convergence results of Theorems 2 and 3 to the network dimensioning problem. In Section 4.3, it is shown that the steady state throughput of the limiting behavior can be calculated from a well-known TCP throughput model with fixed loss probability, *e.g.,* [40, 48].

We now consider a simple application of this limiting result; An ISP currently services up to N_1 TCP flows at peak hour through an ECN/RED access gateway connecting to the core network with the link speed of $N_1 C$ packets/second. The network manager can roughly determine the buffer utilization in the ECN/RED

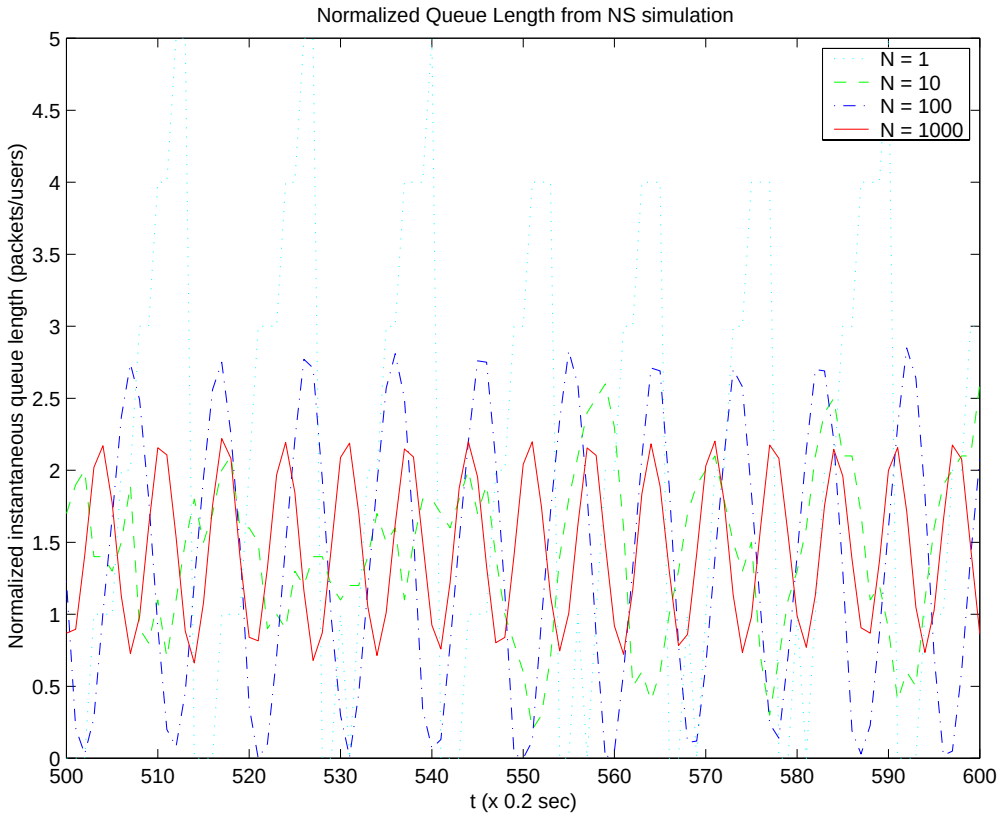

Figure 4.6: The normalized queue length of the ECN/RED gateway in NS simulation with the marking probability function (4.32).

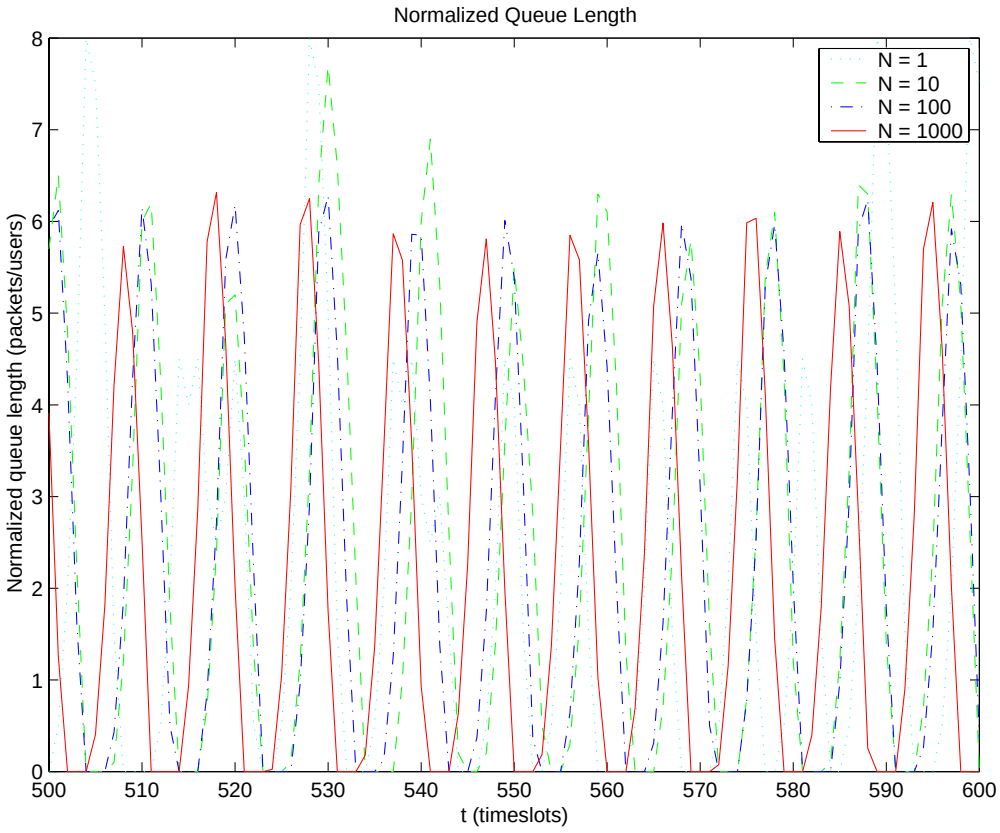

Figure 4.7: The normalized queue length of the model with the marking probability function (4.32).

gateway as follows:

(i) Determine the marking probability per flow ($p = f(q) = 1 - \gamma$) from the relation $C = \mathbf{E}[W^{\gamma}]$ by using a TCP throughput formula such as (4.17);

(ii) Calculate the limiting queue length q in steady state by solving $p = f(q)$;

(iii) Approximate the queue length distribution in steady state via the CLT complement. If the steady state exists, the CLT complement determines the distribution of the queue size fluctuations around q. The delay and overflow distributions can also be approximated via the CLT complement.

While these limiting results apply only to TCP flows with identical round-trip, there are situations where they could be useful. For example, the buffer dimensioning problem in an intercontinental Internet link where it is typically a bottleneck, its large propagation delay dominates the round-trip and the number of flows is extremely large.

Chapter 5

A Window-based Model with Session-level Dynamics and Heterogeneous Round-trip Delays

The main goal of this thesis is in modeling TCP traffic in a realistic situation where there are many flows competing for bandwidth. This is a very challenging problem due to the following factors: (i) the state space explosion, (ii) AQM schemes, (iii) session-layer dynamics, and (iv) variable RTTs of TCP flows. The model presented in Chapter 4 takes into account the factor (i)–(ii). In this chapter, we extend this model to incorporate the session-level dynamics and variable round-trip delays of TCP connections.

Accurate traffic modeling of a large number of TCP flows with session dynamics and heterogeneous RTTs is extremely difficult due to the interactions between session, transport, and network layers, notwithstanding the state space explosion mentioned earlier and the variable feedback information rate. As presented in the literature survey in Section 2.2, ad-hoc assumptions are typically required to make the analysis tractable under a certain regime.

In this chapter, we present a novel model which incorporates not only the interaction of the 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 Chapter 3 and 4, where "macro-scale" modeling of aggregate TCP flows can be developed by systematically applying limit theorems to describe the regime when the number of TCP flows is large.

The chapter is organized as follows. We first present our extended model in Section 5.1. Based on this model, we establish a Weak Law of Large Numbers in Section 5.2. The result shows 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 supports the belief that the RED mechanism indeed helps break the synchronization among the flows suffered by drop-tail gateways. In Section 5.3, we show that the limiting model is also consistent with other previously proposed models in their respective regimes. When the capacity is very small the queue behavior approaches that of a processor sharing queue and when the capacity is very large, its behavior is similar to that of a time-reversed shot-noise model. In addition, we present a simple analysis on the buffer utilization and window size of the limiting model at the steady-state in Section 5.4. Under mild assumptions, the steady-state analysis suggests that only the mean RTT affects the mean queue size at the steady-state. However, the variance of the round-trip delay plays a role in the magnitude of the queue fluctuations. This will be shown with the help of a Central Limit analysis in Section 5.5.

5.1 The Model

Our model captures three layers of dynamics, namely the network, transport, and session layers, which interact with each other through mechanisms to be specified shortly. At the lowest level, the network is simplified to be a single bottleneck router, namely, a RED gateway with ECN capability. The traffic injected into the network is shaped by the TCP congestion-control mechanism in the transport layer, which reacts to the marks from the network. Each TCP connection is initiated by a session, such session being either active or idle. If a session is active, a file or an object is transferred through a TCP connection. An activity period for a session lasts until it no longer has any 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 a detailed description of each of the three layers of the model and of their interactions.

Time is slotted in contiguous timeslots. Here the RTTs of TCP connections are approximated as integer multiples of timeslots, *i.e.,* a timeslot is the greatest common divisor of the RTTs of TCP flows. Similar to the model in Chapter 4, we fix $t = 0, 1, \ldots$ to indicate the start of timeslot $[t, t+1)$.

Heterogeneous round-trip times

Fix $i \in \mathcal{N}$. We assume that 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 timeslot $[t, t+1)$ is denoted by the $\mathcal{D}$-valued rv

$D_i^{(N)}(t)$ with $\mathcal{D} = \{2, 3, \ldots, D_{max}\}$.[1] The bound on the maximum RTT does not constrain the model because actual TCP flows also cannot have larger RTTs than the timeout value. 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

$$\beta_i^{(N)}(t+1) = \left(1 + \beta_i^{(N)}(t)\mathbf{1}\left[\beta_i^{(N)}(t) < D_i^{(N)}(t)\right]\right)\mathbf{1}\left[X_i^{(N)}(t) > 0\right] \tag{5.1}$$

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

Given i in $\mathcal{N}$ and $t = 0, 1, \ldots$, we write

$$G_{i,t}(a,b) = a \cdot \mathbf{1}\left[\beta_i^{(N)}(t) < D_i^{(N)}(t)\right] + b \cdot \mathbf{1}\left[\beta_i^{(N)}(t) \geq D_i^{(N)}(t)\right], \quad a, b \in \mathbb{R} \tag{5.2}$$

in order to simplify our notation later.

Given collections of $\mathbb{R}$-valued rvs $\{Y_i(t), t = 0, 1, \ldots\}$, and $\{Y_{i,new}(t), t = 0, 1, \ldots\}$, we see that

$$\begin{aligned} Y_i(t+1) &= G_{i,t+1}\left(Y_i(t), Y_{i,new}(t+1)\right) \\ &= \begin{cases} Y_{i,new}(t+1), & \beta_i^{(N)}(t+1) \geq D_i^{(N)}(t+1) \\ Y_i(t), & \text{otherwise.} \end{cases} \end{aligned}$$

In other words, the value of $Y_i(t+1)$ is updated to $Y_{i,new}(t+1)$ only at the end of round-trip. Otherwise, the value of $Y_i(t+1)$ remains to be $Y_i(t)$ since no action will be taken before the end of round-trip.

[1] Although $\mathcal{D}$ does not include one in our model, it can be included at the price of more cumbersome proofs. Moreover, this does not cause any loss of generality of the model.

Session dynamics

Each session i in $\mathcal{N}$ is either active or idle. A session idle at the beginning of timeslot $[t, t+1)$ has no packet to transmit in that timeslot. An idle session in timeslot $[t, t+1)$ becomes active at the beginning of timeslot $[t+1, t+2)$ with probability $P_{ar}, 0 < P_{ar} < 1$, independently of past events. In other words, the duration of an idle period is *geometrically* distributed with parameter P_{ar} (hence with mean $1/P_{ar}$). This attempts to capture the dynamics of connection arrivals, where the interarrival times are reported to be exponentially distributed [53].[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 let $\mathbf{1}[U_i(t+1) \leq P_{ar}]$ 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 distributed according to a general probability mass function (pmf) F on $\{1, 2, \ldots\}$. The workload of a connection for session i that becomes active at the beginning of timeslot $[t, t+1)$ is given by $F_i(t)$. This workload represents the *total* number of TCP segments[3] the connection will have to transmit during this activity period. Thus, if a given TCP connection is used to transfer more than one object, this workload variable $F_i(t)$ represents the total number of TCP segments brought in by all these objects. We denote by $X_i(t)$ the remaining workload (expressed in packets) of connection i at the beginning of timeslot $[t, t+1)$. Clearly, we have $X_i(t) = 0$ if session i is idle during $[t, t+1)$.

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

[3]In this model each TCP segment is transmitted as a separate packet.

The evolution of $X_i(t)$ is then given by the recursion

$$\begin{aligned} X_i^{(N)}(t+1) &= \mathbf{1}\left[X_i^{(N)}(t) > 0\right]\left(X_i^{(N)}(t) - A_i^{(N)}(t)\right) \\ &\quad + \mathbf{1}\left[X_i^{(N)}(t) = 0\right]\mathbf{1}\left[U_i(t+1) \le P_{ar}\right]F_i(t+1)\ , \end{aligned} \tag{5.3}$$

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 next 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

$$\begin{aligned} D_i^{(N)}(t+1) = D_i^{(N)}(t)\mathbf{1}\left[X_i^{(N)}(t) > 0\right] \\ + \mathbf{1}\left[X_i^{(N)}(t) = 0\right]\mathbf{1}\left[U_i(t+1) < P_{ar}\right]D_{i,new}(t+1), \end{aligned} \tag{5.4}$$

where the $\mathcal{D}$-valued rvs $\{D_{i,new}(t+1),\ t = 1, 2, \ldots\}$ are i.i.d. rvs which determine the RTT of newly arrived connections.

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 rv $W_i^{(N)}(t)$ has range $\{0, 1, \cdots, W_{\max}\}$ where $W_{\max}$ is a finite integer representing the receiver advertised window size of the TCP connection and the congestion window size of an idle session is taken to be zero. When an idle session becomes active at the beginning of timeslot $[t, t+1)$, the congestion window size of the TCP connection is set to one at the beginning of timeslot $[t+1, t+D_{i,new}(t+1)+1)$, where $D_{i,new}(t+1)$ is the RTT (determined from (5.4)) for the new active session. This models one round-trip delay for the

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

(i) In the Internet, most of the packet arrivals at a bottleneck are usually compressed together due to the "ACK compression" phenomenon [64], 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 a 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 upperbound 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. Therefore, the difference in the control action due to our bursty packet arrivals will be smoothed out by the averaging mechanism with long memory in RED.

Suppose that connection i has $X_i^{(N)}(t)$ remaining packets (or workload) waiting to be transmitted at the beginning of timeslot $[t, t+1)$,[4] the number of

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

packets transmitted by connection i at the beginning of timeslot $[t, t+1)$, denoted by $A_i^{(N)}(t)$, is given by

$$A_i^{(N)}(t) = \min\left(W_i^{(N)}(t), X_i^{(N)}(t)\right) \mathbf{1}\left[\beta_i^{(N)}(t) \geq D_i^{(N)}(t)\right]. \tag{5.5}$$

Note from (5.5) that a connection transmits only once per RTT.

The congestion-control mechanism of TCP operates either in the *slow start* (SS) or the *congestion avoidance* (CA) phase. A new TCP connection starts in SS in order to quickly gauge the available bandwidth of the network. While in SS, the congestion window size is doubled every round-trip time 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, if the connection is in SS, then the congestion window of connection i evolves according to

$$\begin{aligned} W_{i,SS}^{(N)}(t+1) &= \min\left(2W_i^{(N)}(t) \vee 1, W_{\max}\right) M_i^{(N)}(t+1) \\ &\quad + \lceil \frac{W_i^{(N)}(t)}{2} \rceil \left(1 - M_i^{(N)}(t+1)\right), \end{aligned} \tag{5.6}$$

where $M_i^{(N)}(t+1)$ is an indicator function of the event that no packet of connection i has been marked in the round-trip preceding timeslot $[t, t+1)$, i.e., $M_i^{(N)}(t+1) = 1$ when no packet from session i is marked and $M_i^{(N)}(t+1) = 0$ when at least one packet is marked. The marking mechanism is explained in the next subsection.

In CA, the congestion window size in the next timeslot $[t+1, t+2)$ is increased by one if no mark is received in timeslot $[t, t+1)$, while if one or more packets are marked in timeslot $[t, t+1)$, the congestion window in the next

timeslot is reduced by half. The congestion window size in CA is then given by

$$W_{i,CA}^{(N)}(t+1) = \min\left(W_i^{(N)}(t)+1, W_{\max}\right) M_i^{(N)}(t+1) + \lceil \frac{W_i^{(N)}(t)}{2} \rceil \left(1 - M_i^{(N)}(t+1)\right). \quad (5.7)$$

Both $W_{i,SS}^{(N)}(t+1)$ and $W_{i,CA}^{(N)}(t+1)$ are candidates for the congestion window size in timeslot $[t+1, t+2)$ depending on the state that the connection i is in.

Since we only update the congestion window at the end of the round-trip, we use the mapping (5.2) to retain the value of the congestion window until the end of the round-trip where it is updated. If connection i is in SS in timeslot $[t, t+1)$, its potential window $\hat{W}_{i,SS}^{(N)}(t+1)$ in timeslot $[t+1, t+2)$ is given by

$$\hat{W}_{i,SS}^{(N)}(t+1) = G_{i,t+1}\left[W_i^{(N)}(t), W_{i,SS}^{(N)}(t+1)\right], \quad (5.8)$$

where $W_{i,SS}^{(N)}(t+1)$ is given in (5.6). Similarly, if connection i is in CA in timeslot $[t, t+1)$, its potential window $\hat{W}_{i,CA}^{(N)}(t+1)$ in timeslot $[t+1, t+2)$ is given by

$$\hat{W}_{i,CA}^{(N)}(t+1) = G_{i,t+1}\left[W_i^{(N)}(t), W_{i,CA}^{(N)}(t+1)\right], \quad (5.9)$$

where $W_{i,CA}^{(N)}(t+1)$ is given in (5.7).

Let the $\{0,1\}$-valued rvs $\{S_i^{(N)}(t), i \in \mathcal{N}\}$ encode the state of TCP connections, with the interpretation that $S_i^{(N)}(t) = 0$ (resp. $S_i^{(N)}(t) = 1$) if connection i is in CA (resp. in SS) at the beginning of timeslot $[t, t+1)$. Therefore, combining (5.8) and (5.9), we see that the complete recursion of the congestion window size can be written as

$$W_i^{(N)}(t+1) = \mathbf{1}\left[X_i^{(N)}(t) - A_i^{(N)}(t) > 0\right] \times \left(S_i^{(N)}(t)\hat{W}_{i,SS}^{(N)}(t+1) + (1 - S_i^{(N)}(t))\hat{W}_{i,CA}^{(N)}(t+1)\right). \quad (5.10)$$

The first indicator function in (5.10) 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 $\{S_i^{(N)}(t),\ t = 0, 1, \ldots\}$ is given by

$$\begin{aligned} S_i^{(N)}(t+1) &= \mathbf{1}\left[X_i^{(N)}(t) \leq W_i^{(N)}(t)\right] \\ &\quad +\mathbf{1}\left[X_i^{(N)}(t) > W_i^{(N)}(t)\right] S_i^{(N)}(t) M_i^{(N)}(t+1). \end{aligned} \tag{5.11}$$

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. Equation (5.11) assumes that a new TCP connection in SS is ready to be set up one timeslot after the previous connection is torn down upon 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.

Network dynamics

The model of the network dynamics is identical to the model presented earlier in Chapter 4 with two exceptions. First, we also introduce a queue averaging mechanism to further generalize the model. The marking probability will be a function of the average queue size as is implemented in RED instead of the instantaneous queue size. Second, the indicator function $M_i^{(N)}(t+1)$ needs to be modified in order to represent the event that no marks have been received in the

previous round-trip (as opposed to the previous timeslot as in Chapter 4). The remaining dynamics are the same as will be briefly summarized here.

The recursion of the queue is written in the a similar to Lindley's recursion introduced in (1.9). More specifically, it can be rewritten as

$$\begin{aligned} & Q^{(N)}(t+1) \\ = & \left[Q^{(N)}(t) - NC + \sum_{i=1}^{N} \min\left(W_i^{(N)}(t), X_i^{(N)}(t) \right) \mathbf{1}\left[\beta_i^{(N)}(t) \geq D_i^{(N)}(t) \right] \right]^+, \end{aligned}$$

where $A^{(N)}(t) = \sum_{i=1}^{N} A_i^{(N)}(t)$ and $A_i^{(N)}(t)$ is given as (5.5) in this model.

The queue averaging mechanism utilizes an exponentially weighted moving average (EWMA) filter to average the instantaneous queue size with the time constant depending on the parameter $0 < \alpha \leq 1$. Let $\hat{Q}^{(N)}(t+1)$ represent the EWMA queue in the beginning of timeslot $[t+1, t+2)$, then

$$\hat{Q}^{(N)}(t+1) = (1-\alpha)\hat{Q}^{(N)}(t) + \alpha Q^{(N)}(t+1). \tag{5.12}$$

A special case when $\alpha = 1$ is equivalent to using the instantaneous queue size for marking mechanism as in Chapter 4.

We represent the marking through the $\{0,1\}$-valued rvs $M_{i,j}^{(N)}(t+1)$ $(j = 1, ..., A_i^{(N)}(t))$ with $M_{i,j}^{(N)}(t+1) = 0$ (resp. $M_{i,j}^{(N)}(t+1) = 1$) if the jth packet from source i is marked (resp. not marked) in the RED buffer. More concretely, for each i in $\mathcal{N}$ and $j = 1, 2, \ldots$, we write

$$M_{i,j}^{(N)}(t+1) = \mathbf{1}\left[V_{i,j}(t+1) > f^{(N)}(\hat{Q}^{(N)}(t)) \right], \tag{5.13}$$

where again the collection of i.i.d. $[0,1]$-*uniform* rvs $\{V_{i,j}(t+1), V_i(t+1),\ i, j = 1, \cdots;\ t = 0, 1, \cdots\}$ are assumed independent of all other rvs introduced so far.

The indicator function of the event that no packets from connection i in timeslot $[t, t+1)$ are marked can then be written as

$$M_{i,new}^{(N)}(t+1) = \begin{cases} \prod_{j=1}^{A_i^{(N)}(t)} M_{i,j}^{(N)}(t+1), & A_i^{(N)}(t) \geq 1 \\ 1, & A_i^{(N)}(t) = 0. \end{cases} \tag{5.14}$$

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

$$M_i^{(N)}(t+1) = G_{i,t}(M_i^{(N)}(t), M_{i,new}^{(N)}(t+1)). \tag{5.15}$$

Notice that we replace $t+1$ in the mapping G in (5.2) by t to delay the update in the value of $M_i^{(N)}(t+1)$ in order for (5.8) and (5.9) to evolve based on the markings in the previous round-trip. For example, if $W_i^{(N)}(t)$ is updated in timeslot $[t, t+1)$ then $M_i^{(N)}(t+1)$ is updated in timeslot $[t+1, t+2)$ and its value will be used in the timeslot $[t + D_i^{(N)}(t), t + D_i^{(N)}(t) + 1)$ to determine the new congestion window size.

5.2 The Asymptotics

The first main result of the chapter consists of the asymptotics for the normalized buffer content as the number of sessions becomes large similar to Theorem 2. This result is again discussed under the following Assumptions (AW1),(AW2b) where the structural condition (AW1) is identical to that in Chapter 4 while (AW2b) is modified from (AW2) to incorporate additional initial conditions, *i.e.,*

(AW2b) For each $N = 1, 2, \ldots$ and $i = 1, \ldots, N$, the initial conditions of rvs in the model are given by

$$Q^{(N)}(0) = \hat{Q}^{(N)}(0) = W_i^{(N)}(0) = \beta_i^{(N)}(0) = D_i^{(N)}(0) = 0,$$

and

$$S_i^{(N)}(0) = M_i^{(N)}(0) = 1.$$

We denote the vector of state variables for session i in timeslot $[t, t+1)$ by

$$\mathbf{Y}_i^{(N)}(t) := (W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), D_i^{(N)}(t), \beta_i^{(N)}(t), M_i^{(N)}(t)). \tag{5.16}$$

The rv $\mathbf{Y}_i^{(N)}(t)$ takes a value in the discrete set

$$\begin{aligned} \mathcal{Y} \;:=\; & \{0,1,\ldots,W_{max}\} \times \{0,1,\ldots,X_{max}\} \times \{0,1\} \times \{0,2,3,\ldots,D_{max}\} \\ & \times \{0,1,\ldots,D_{max}\} \times \{0,1\}. \end{aligned} \tag{5.17}$$

Theorem 4. *Assume that (AW1) and (AW2b) hold. Then, for each $N = 1, 2, \ldots$ and $t = 0, 1, \ldots$, there exists a (non-random) constant $q(t)$ and a $\mathcal{Y}$-valued 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:

$$\frac{Q^{(N)}(t)}{N} \;\stackrel{P}{\to}_N\; q(t), \quad \frac{\hat{Q}^{(N)}(t)}{N} \stackrel{P}{\to}_N \hat{q}(t) \tag{5.18}$$

and

$$\mathbf{Y}_1^{(N)}(t) \Longrightarrow_N \mathbf{Y}(t). \tag{5.19}$$

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

$$\tfrac{1}{N}\textstyle\sum_{i=1}^N g\left(\mathbf{Y}(t)\right) \stackrel{P}{\to}_N \mathbf{E}\left[g\left(\mathbf{Y}(t)\right)\right]. \tag{5.20}$$

(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

$$\lim_{N\to\infty} \mathbf{P}[\mathbf{Y}_i^{(N)}(t) = \mathbf{y}_i,\ i = 1, \cdots, I] = \prod_{i=1}^{I} \mathbf{P}\left[\mathbf{Y}(t) = \mathbf{y}_i\right] \tag{5.21}$$

for any $\mathbf{y}_i \in \mathcal{Y}$, $i = 1, \ldots, I$.

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

$$q(t+1) = (q(t) - C + \mathbf{E}\left[A(t)\right])^+ \tag{5.22}$$

and

$$\hat{q}(t+1) = (1-\alpha)\hat{q}(t) + \alpha q(t+1) \tag{5.23}$$

where

$$A(t) = \min\left(W(t), X(t)\right) \mathbf{1}\left[\beta(t) \geq D(t)\right].$$

Further, the recurrence

$$\begin{aligned} \mathbf{Y}(t+1) &= (W(t+1), X(t+1), S(t+1), D(t+1), \beta(t+1), M(t+1)) \\ &=_{st} P(\mathbf{Y}(t)) \\ &:= (P_1(\mathbf{Y}(t)), P_2(\mathbf{Y}(t)), \ldots, P_6(\mathbf{Y}(t))) \end{aligned} \tag{5.24}$$

holds in law, where $P_1(\mathbf{Y}(t)), \ldots, P_6(\mathbf{Y}(t))$ *are given by*

$$\begin{aligned}
P_1(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) - A(t) > 0\right]\left(S(t)W_{SS}(t+1) + (1 - S(t))W_{CA}(t+1)\right), \\
P_2(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) > 0\right]\left(X(t) - A(t)\right) + \mathbf{1}\left[X(t) = 0\right]\mathbf{1}\left[U(t+1) < P_{ar}\right]F(t+1), \\
P_3(\mathbf{Y}(t)) &= \mathbf{1}\left[X(t) - A(t) \leq 0\right] + \mathbf{1}\left[X(t) - A(t) > 0\right]S(t)M(t+1), \\
P_4(\mathbf{Y}(t)) &= D(t)\mathbf{1}\left[X(t) > 0\right] + \mathbf{1}\left[X(t) = 0\right]\mathbf{1}\left[U(t+1) < P_{ar}\right]D(t+1), \\
P_5(\mathbf{Y}(t)) &= \left(1 + \beta(t)\mathbf{1}\left[\beta(t) < D(t)\right]\right)\mathbf{1}\left[X(t) > 0\right],
\end{aligned}$$

and

$$P_6(\mathbf{Y}(t)) = G_t(M(t), \mathbf{1}\left[V(t+1) \leq (1 - f(\hat{q}(t)))^{A(t)}\right]),$$

where

$$\begin{aligned}
M(t+1) &:= P_6(\mathbf{Y}(t)), \\
G_t(a,b) &= a\mathbf{1}\left[\beta(t) < D(t)\right] + b\mathbf{1}\left[\beta(t) \geq D(t)\right], \quad a, b \in \mathbb{R}, \\
W_{SS}(t+1) &= G_{t+1}\left(W(t), \min\left(2W(t) \vee 1, W_{\max}\right)M(t+1) + \lceil\frac{W(t)}{2}\rceil(1 - M(t+1))\right),
\end{aligned}$$

and

$$W_{CA}(t+1) = G_{t+1}\left(W(t), \min\left(W(t)+1, W_{\max}\right)M(t+1) + \lceil\frac{W(t)}{2}\rceil(1 - M(t+1))\right),$$

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

The proof of Theorem 4 is presented in Chapter 6.

5.3 Limiting Cases

Next, we briefly consider the resulting model from Theorem 4 in the regime when C is either very large or very small under the following assumption:

(AW3) The marking function $f : \mathbb{R} \to [0,1]$ is monotonically increasing with $f(0) = 0$ and $\lim_{x\to\infty} f(x) = 1$;

It is easy to see that $\lim_{C\to\infty} q(t) = 0$ for all $t = 0, 1, \ldots$, so that the marking probability per flow also converges to zero from (AW3) for all t. Therefore, each incoming flow will always operate in the slow start (exponential growth) mode and the resulting input traffic into the network is the superposition of (discrete-time) Poisson arrival streams of random number of packets, each of which doubles its window size every round-trip. The aggregate input traffic is therefore similar to the time-reversed shot-noise processes, which is compatible with the model studied in [23].

On the other hand, with $C \simeq 0$, the queue will start building up, whence $\lim_{t\to\infty} q(t) = \infty$. Thus, for large t, all TCP flows (including incoming TCP flows) will experience marking probability close to one under Assumption (AW3). All active connections will be able to inject only one packet per round-trip into the network because every packet transmitted will be marked with a probability going to one. As a result, each TCP congestion window size approaches one with high probability. Since the bottleneck router will transmit packets non-selectively, any active flow will receive roughly equal throughput and hence the queue behavior approaches that of processor-sharing, assuming identical RTTs among all flows, which is in agreement with [32].

5.4 Steady-State Regime

Using the results from the previous sections, we now carry out a simple steady-state analysis with a few additional reasonable assumptions. Although we have not formally proved that the system converges to steady state as complications arise due the deterministic (*i.e.,* degenerate) character of the first

two components in the time-homogeneous Markov chain $\{(q(t), \hat{q}(t), \mathbf{Y}(t)),\ t = 0, 1, \ldots\}$, simulation results indicate that the system does converge to steady state for a large range of initial conditions and parameters. This is illustrated in Section 5.6. In order to facilitate our further analysis we assume the following:

(AW4b) The sequence $\{(q(t), \mathbf{Y}(t)),\ t = 0, 1, \ldots\}$ admits a steady state in the sense that $(q(t), \mathbf{Y}(t)) \Longrightarrow_t (q^\star, \mathbf{Y}^\star)$ 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 $\mathcal{Y}$-valued rv.

(AW5) Let F_{ar} be a rv with the distribution F, representing the initial workload size of a new connection. we assume $\mathbf{E}[W^\star] \ll \mathbf{E}[F_{ar}]$.

(AW6) We assume that when an active connection finishes its last transmission, it waits an additional RTT before resetting its window size to zero. [5]

We first introduce a simple result which will help in the analysis. Its proof is given in Section 5.7.

Lemma 1. *Assuming (AW1),(AW2b),(AW3), (AW4b),(AW5)-(AW6), the rvs* $\min(W^\star, X^\star)$ *and* $\mathbf{1}[\beta^\star \geq D^\star]$ *are conditionally independent given that the connection is active.*

It is easily seen that Assumption (AW4b) immediately implies $\hat{q}(t) \rightarrow_t \hat{q}^\star = q^\star$ for some constant $\hat{q}^\star$. And so 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

[5] This will have only a marginal effect under Assumption (AW5).

solution to

$$q^\star = (q^\star - C + \mathbf{E}\left[A^\star\right])^+$$
$$= (q^\star - C + \mathbf{E}\left[\min(W^\star, X^\star)\mathbf{1}\left[\beta^\star \geq D^\star\right]\right])^+. \tag{5.25}$$

Since the window size and the workload are both zero when a session is idle, we have

$$\begin{aligned}\mathbf{E}\left[A^\star\right] &= \mathbf{P}\left[\text{active}\right]\mathbf{E}\left[A^\star|\text{active}\right] \\ &= \mathbf{P}\left[\text{active}\right]\mathbf{E}\left[\min(W^\star, X^\star)|\text{active}\right]\mathbf{P}\left[\beta^\star \geq D^\star|\text{active}\right] \end{aligned} \tag{5.26}$$

where the last equality follows from Lemma 1. The probability $\mathbf{P}\left[\text{active}\right]$ that a session is active in steady state is given by

$$\mathbf{P}\left[\text{active}\right] = \frac{\mathbf{E}\left[\text{connection duration}\right]}{\mathbf{E}\left[\text{connection duration}\right] + \mathbf{E}\left[\text{idle period}\right]}$$

by elementary arguments from renewal theory.

First, we can rewrite
$\mathbf{P}\left[\beta^\star \geq D^\star|\text{active}\right] = \sum_{d_i}\mathbf{P}\left[\beta^\star \geq d_i|\text{active}, D^\star = d_i\right]\mathbf{P}\left[D^\star = d_i|\text{active}\right]$.
Conditioning on the event that $D^\star = d$, it is easy to see that

$$\mathbf{P}\left[D^\star = d_i|\text{active}\right] \approx \frac{d_i\mathbf{P}\left[D = d_i\right]}{\sum_{d_j\in\mathcal{D}} d_j\mathbf{P}\left[D = d_j\right]}.$$

The expression is not exact because according to our model, the connection resets its window to zero one timeslot after transmitting its last packet. However, this will have only a marginal effect under (AW5).

From Assumption (AW5) since a connection typically lasts many RTTs, we have

$$\mathbf{P}\left[\beta^\star \geq d_i|\text{active}, D^\star = d_i\right] \approx \frac{1}{d_i}. \tag{5.27}$$

Therefore,

$$\begin{aligned}\mathbf{P}\left[\beta^\star \geq D^\star|\text{active}\right] &\approx \sum_{d_i\in\mathcal{D}} \frac{\mathbf{P}\left[D=d_i\right]}{\sum_{d_j\in\mathcal{D}} d_j\mathbf{P}\left[D=d_j\right]} \\ &= \frac{1}{\mathbf{E}\left[D\right]}. \end{aligned} \tag{5.28}$$

Let F_{ar} be the initial workload and D be the RTT of a connection. The conditional expected value of the connection duration given the workload size and round-trip delay is

$$\mathbf{E}\left[\text{connection duration} \mid F_{ar}=x, D=d\right] = \frac{x}{T(d, f(q^\star))} \ ,$$

where $T(d, f(q^\star))$ is the mean throughput of a TCP connection with RTT of d and packet marking probability of $f(q^\star)$. Here we approximate the average throughput of a TCP connection by the well-known throughput formula

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

where K is some constant in the interval $[1, \frac{8}{3}]$, according to [48, 40].

This implies that

$$\mathbf{E}\left[\text{connection duration}|F_{ar}=x, D=d\right] \approx \frac{xd\sqrt{f(q^\star)}}{K}.$$

Since the initial workload and the RTT are assumed independent, we have $\mathbf{E}\left[\text{connection duration}\right] \approx \frac{\mathbf{E}[F_{ar}]\mathbf{E}[D]\sqrt{f(q^\star)}}{K}$. Therefore,

$$\mathbf{P}\left[\text{active}\right] \approx \frac{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[D\right]\sqrt{f(q^\star)}}{\mathbf{E}\left[F_{ar}\right]\mathbf{E}\left[D\right]\sqrt{f(q^\star)} + K/P_{ar}} \tag{5.29}$$

Finally, in order to compute (5.26), 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

(AW5), *i.e.,* $\mathbf{E}[F] >> \mathbf{E}[W^\star]$.

$$\begin{aligned}\mathbf{E}[W^\star|\text{active}] &= \sum_{d_i\in\mathcal{D}} \mathbf{E}[W^\star|D^\star = d_i, \text{active}]\,\mathbf{P}[D^\star = d_i|\text{active}] \\ &= \sum_{d_i\in\mathcal{D}} \frac{K}{\sqrt{f(q^\star)}} \frac{d_i\mathbf{P}[D=d_i]}{\sum_{d_j\in\mathcal{D}} d_j\mathbf{P}[D=d_j]} \\ &= \frac{K}{\sqrt{f(q^\star)}}. \end{aligned} \tag{5.30}$$

Combining (5.26)-(5.30), we get

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

If $0 < f(q^\star) < 1$, it is necessary that $C = \mathbf{E}[A^\star]$. After some simple algebras, we can show

$$f(q^\star) \approx \frac{K^2}{\mathbf{E}[D]^2}\left(\frac{1}{C} - \frac{1}{P_{ar}\mathbf{E}[F_{ar}]}\right)^2. \tag{5.31}$$

If f is invertible, then

$$q^\star \approx f^{-1}\left(\frac{K^2}{\mathbf{E}[D]^2}\left(\frac{1}{C} - \frac{1}{P_{ar}\mathbf{E}[F_{ar}]}\right)^2\right). \tag{5.32}$$

Note that the steady-state marking probability and the average queue size depend on the round-trip delay and incoming workload distributions only through their *mean values*. Numerical examples validating the analysis are given in Section 5.6. 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, given the system parameters. However, 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 will be shown in the Central Limit analysis in the next section.

5.5 A Central Limit Theorem

In this section we present a CLT complement similar to Theorem 3 to the LLN results in Theorem 4. The description of the model and the notations are given in Section 5.1. The analysis is carried out under the same model and we again need to strengthen Assumption (AW1) to (AW1b) as given in Section 4.4. However, we also introduce Assumption (A7):
(AW7) The workload of a new TCP connection is bounded, *i.e.,* there exists an integer X_{max} such that $F_i^{(N)}(t)$ is a $\{1, \cdots, X_{max}\}$-valued rv for all i in $\mathcal{N}$ and all $t = 0, 1, \ldots$. [6]

Fix $t = 0, 1, \ldots$. Again, the asymptotic residual capacity given in (4.27) plays a crucial role in the analysis. In this model, it is given by

$$K(t) \;=\; C - q(t) - \mathbf{E}\left[\min(W(t), X(t))\mathbf{1}\left[\beta(t) \geq D(t)\right]\right].$$

Now define a collection of rvs that is integral to the analysis. For each $N = 1, 2, 3, \ldots$ and $\mathbf{y} = (w, x, s, d, b, m) \in \mathcal{Y}$, let $L_0^{(N)}(t)$ be as given in (4.23). We also introduce the following notations:

$$\hat{L}_0^{(N)}(t) \;:=\; \frac{\hat{Q}^{(N)}(t)}{N} - \hat{q}(t), \tag{5.33}$$

and

$$L_{\mathbf{y}}^{(N)}(t) := \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \mathbf{1}_{\mathbf{Y}_i^{(N)}(t)}\left[\mathbf{y}\right] - \mathbf{P}_{\mathbf{Y}(t)}\left[\mathbf{y}\right]. \tag{5.34}$$

Theorem 5. *Assume (AW1b),(AW2b) and (AW7) hold. Then, for each* $t = 0, 1, \ldots$, *there exists an* $\mathbb{R}^{|\mathcal{Y}|+2}$*-valued rv* $\mathbf{L}(t) = (L_0(t), \hat{L}_0(t), L_{\mathbf{y}}(t), \mathbf{y} \in \mathcal{Y})$ *such*

[6] The limit X_{max} can be lifted. This, however, results in a more complicated analysis. Nevertheless, such a restriction is necessary for the numerical calculation of the CLT on a computer.

that the convergence

$$\sqrt{N}\left(L_0^{(N)}(t), \hat{L}_0^{(N)}(t), L_{\mathbf{y}}^{(N)}(t),\ \mathbf{y} \in \mathcal{Y}\right) \Longrightarrow_N \mathbf{L}(t) \tag{5.35}$$

holds. Moreover, the distributional recurrences

$$L_0(t+1) =_{st} \begin{cases} 0 & K(t) > 0 \\ L_0(t) + \bar{L}(t) & K(t) < 0 \\ \left(L_0(t) + \bar{L}(t)\right)^+ & K(t) = 0 \end{cases} \tag{5.36}$$

and

$$\hat{L}_0(t+1) =_{st} (1-\alpha)\hat{L}_0(t) + \alpha L_0(t+1) \tag{5.37}$$

hold, where

$$\bar{L}(t) = \sum_{\mathbf{y} \in \mathcal{Y}} \min(w, x)\mathbf{1}\left[b \geq d\right] L_{\mathbf{y}}(t).$$

The distribution of the rv $L_{\mathbf{y}}(t),\ \mathbf{y} \in \mathcal{Y},\ t = 0, 1, \ldots$, *can be calculated recursively starting with* $t = 0$.

Finally, for any $t = 1, 2, \ldots$, *the rv* $L_0(t+1)$ *is Gaussian*[7] *if* $K(s) \neq 0$ *for all* $s < t$.

A proof of Theorem 5 and the complete distributional recursions are provided in [57].

From the last statement of Theorem 5, a necessary condition for $L_0(t)$ not to be Gaussian is $K(s) = 0$ for some $s < t$. This is, however, a technical artifact of the limiting regime as in practice it is unlikely that the real-valued residual capacity would attain the value zero *exactly*. Therefore, in practice the distribution of the RED buffer at any fixed time t can be well approximated by a Gaussian rv.

[7]Here we interpret a constant as a Gaussian rv with standard deviation equals to zero.

Discussion

The CLT analysis reveals the sources of fluctuation in the queue size. It is shown in the proof of Theorem 5 that the queue fluctuation $L_0(t+1)$ consists of four components :

(i) Fluctuation caused by the discrepancy between the feedback information from RED to TCP sources $f^{(N)}(\hat{Q}^{(N)}(t))$ and the limiting feedback information $f(\hat{q}(t))$: This uncertainty in feedback information can be explained by the Delta Method (Proposition 1) and is already discussed in details in Section 4.4.

(ii) Binary nature of feedback information: The RED gateway either marks a packet or does not in a RTT. This binary nature of feedback information imposes a limited feedback information granularity, and causes a fluctuation in queue size. This fluctuation can be well approximated by a Gaussian rv.

(iii) Fluctuation caused by the arrival of new TCP connections and the random idle periods: The larger the file size, *i.e.,* workload of a new TCP connection, and waiting time variances, the larger the magnitude of this fluctuation is. This part of the fluctuation can also be described by a Gaussian rv.

(iv) Fluctuation caused by the structure of protocols: The structure of the protocols determines the mappings which combined the rvs discussed in (i)-(iii). The resulting rv represents the overall fluctuation observed at the queue.

Components (ii) and (iv) are due to the protocols and cannot be mitigated without modifying the protocols. Component (iii) depends on users behavior, and hence is beyond the control of the network. Thus, network designers can only

manipulate the shape (or slope) of the feedback function to reduce oscillation of queue size. Although reducing the slope of the queue can decrease the magnitude of fluctuation, it also increases the average queue size as suggested by (5.32).

The aforementioned trade-off can be explored in the context of an optimization problem. Suppose that we have the convergences $q(t) \rightarrow_t q^\star$ and $L_0(t) \Rightarrow_t L_0^\star$.[8] Then, the steady-state queue distribution can be approximated by $Nq^* + \sqrt{N}L_0^\star$. For best-effort traffic, the marking function f should, for example, maximize $\mathbf{P}\left[0 < Nq^* + \sqrt{N}L_0^\star < NB\right]$, assuming that the buffer size is NB. Clearly, a more sophisticated performance metric can be adopted, which depends on the probabilities of an empty queue and of buffer overflow. The availability of the queue distribution allows us to formulate the corresponding optimization problem and to propose a systematic solution.

Furthermore, if the protocol suite (*i.e.,* TCP and ECN/RED) can be modified, then we can further reduce the magnitude of queue fluctuation caused by limited feedback granularity, *i.e.,* component (ii). One simple scheme to improve the feedback information granularity is to increase the number of feedback information bits in the ECN mechanism. Given that the improved feedback information is properly utilized, the magnitude of queue fluctuation would be reduced. Multi-level ECN (MECN) [10] is an example of such a scheme.

[8]Note that both $q^\star$ and $L_0^\star$ depend on f.

5.6 Simulations of the Model with Session-level Dynamics and Variable Round-trip Delays

In the previous chapter, we have presented simulation results which point to the convergence of the normalized queue when the number of persistent TCP flows is large. In this section, we extend the simulations to incorporate both session-level dynamics and the variable round-trip delays of connections in order to demonstrate that a similar convergence still exists and also reveals the roles that file size and round-trip have on the convergence of the queue along with validation of the steady-state asymptotic queue formula in (5.32).

Numerical examples

This section presents numerical examples to study the behavior of the queue size per flow. The numerical examples are evaluated through Monte-Carlo simulations of the model in Section 5.1.

Example (i)

The system and control parameters are set as follows: $C = 1$ packet/timeslot and

$$f^{(N)}(x) = \begin{cases} 0, & x < 2N \\ 0.2\frac{x-2N}{18N}, & 2N \le x \le 20N \\ 1, & \text{otherwise.} \end{cases}$$

The initial values are set to according to Assumption (AW2b). The variables evolve from their initial values according to the dynamics outlined in Section 5.1. The workload $F_i(t) \sim \mathit{geometric}(p),\ i = 1, 2, \ldots, N$ and $t = 0, 1, \ldots$ where

$p = 0.001$, *i.e.,* $\mathbf{E}[F_i(t)] = 1,000$ 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 averaging parameter α is set to 0.01.

First, we simulate the dynamics when the random round-trip delay $D_{i,new}(t)$ is uniformly distributed on the set $\mathcal{D} = \{2, \ldots, 6\}$. Figure 5.1 plots the evolution of the queue size per flow 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 $K = \sqrt{3/2}$ (which is shown in Section 4.3.2 to be a reasonable approximation), we have from (5.32) $q^\star = f^{-1}\left(\frac{3/2}{4^2}\left(\frac{1}{1} - \frac{1}{0.05 \cdot 1000}\right)^2\right) = 10.1$. Hence, as can be seen from Figure 5.1, the steady-state queue size is close to the value predicted by (5.32).

In the next example, we demonstrate the role of the variance of the round-trip delay. In this case, the round-trip delay of each connection is either 2 or 6 with equal probability, *i.e.,* Bernoulli rv with
$\mathbf{P}[D_{i,new}(t) = 2] = \mathbf{P}[D_{i,new}(t) = 6] = 0.5, \quad t = 0, 1, \ldots$, which has the largest variance of any distribution on $\mathcal{D}$ with the mean of 4. The rest of setup is identical to the previous case. Figure 5.2 plots the evolution of the queue size per flow. 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 users 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 depends only on the mean of the round-trip delay, the transient behavior and the magnitude of fluctuation are affected by the variance of round-trip delay.

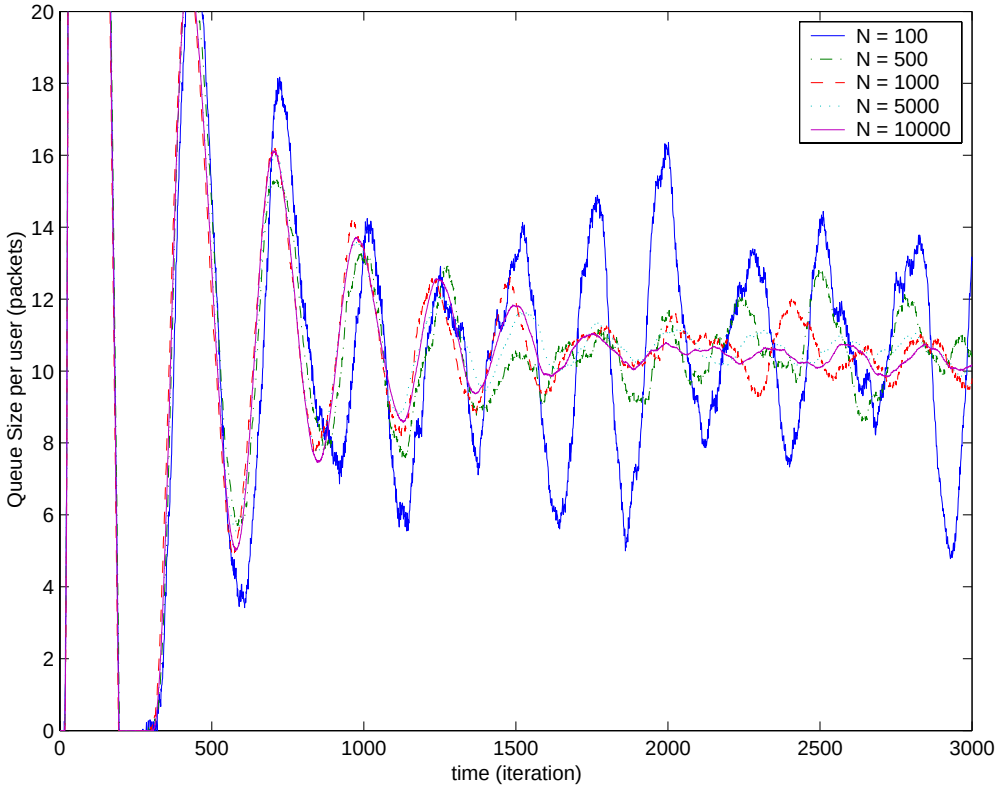

Figure 5.1: Evolution of queue size per flow when the round-trip is uniform.

Example (ii)

In the second example, we change the capacity per flow, workload and the idle period distribution to be as follows: $C = 0.6$ packet/timeslot, and workload $F_i(t) \sim geometric(p), i = 1, \ldots, N$ and $t = 0, 1, \cdots$, 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 $D_{i,new}(t)$ is uniformly distributed on the set $\mathcal{D} = \{2, \ldots, 10\}$ so the average round-trip delay equals 6 timeslots. Figure 5.3 plots the evolution of the queue size per flow

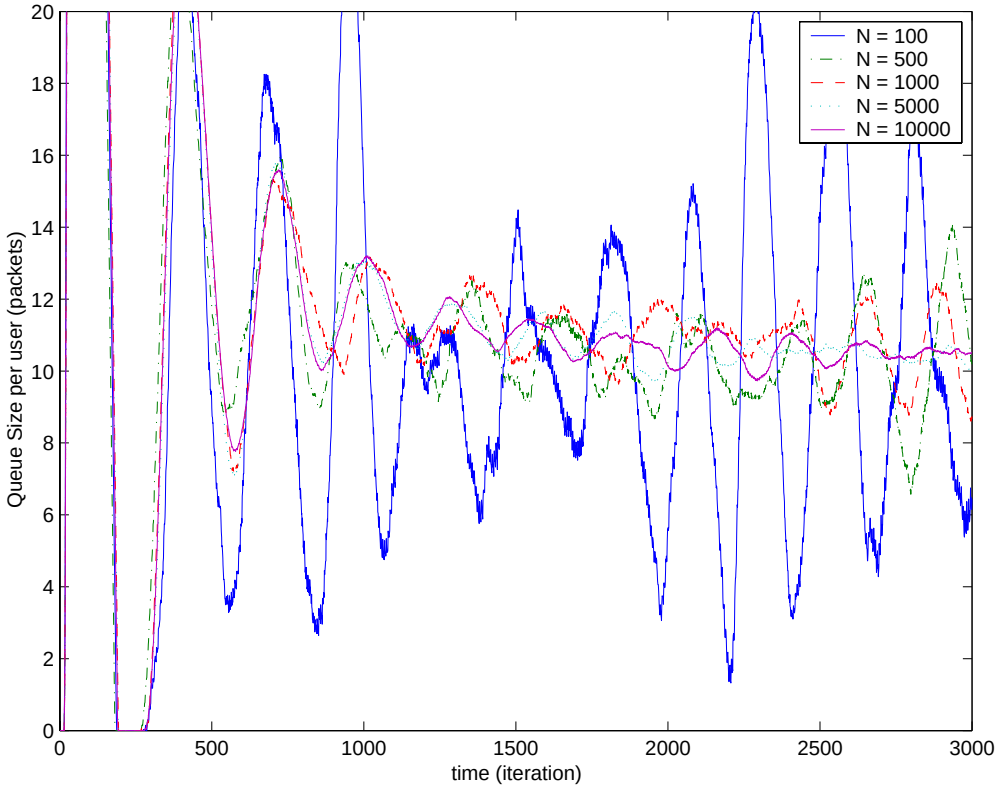

Figure 5.2: Evolution of queue size per flow when the round-trip is Bernoulli.

with the number of sessions $N = 100, 500, 1000, 5000$, and 10000. Figure 5.4 plots the evolution when $D_{i,new}(t)$ has the distribution

$\mathbf{P}\left[D_{i,new}(t) = 2\right] = \mathbf{P}\left[D_{i,new}(t) = 10\right] = 0.5$, $i = 1, 2, \ldots, N$ and $t = 0, 1, \ldots$, so that $\mathbf{E}\left[D_{i,new}(t)\right] = 6$.

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

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

which is close to the steady-state queue level in both simulations. Therefore, this

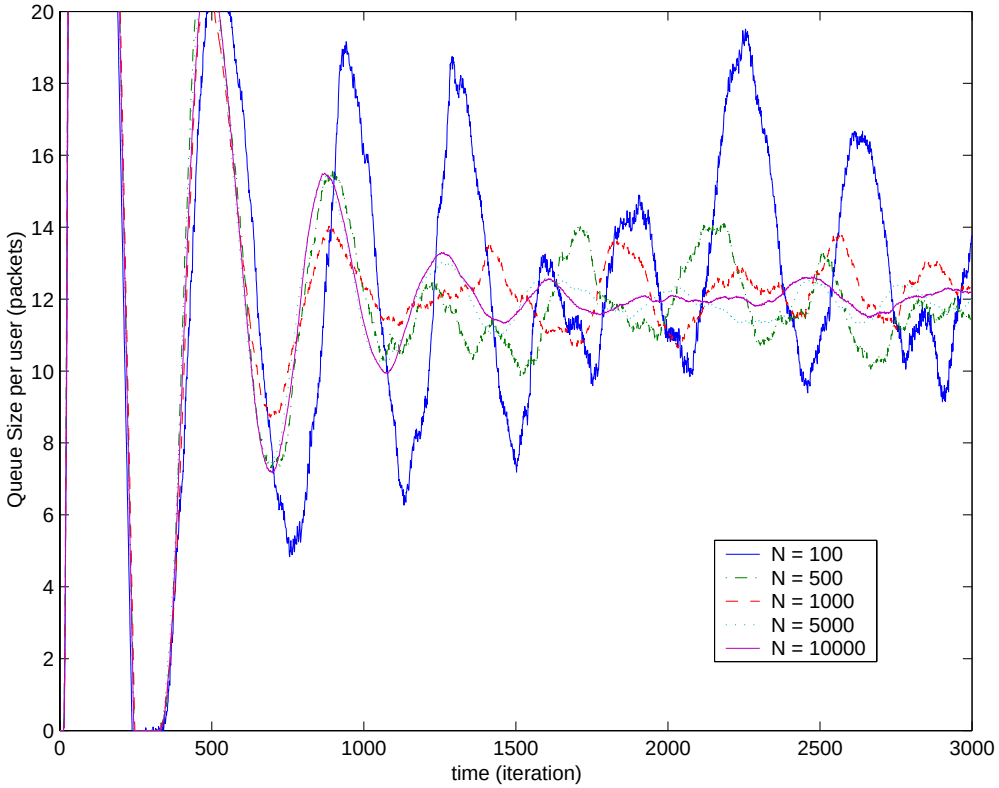

Figure 5.3: Example (ii) : Evolution of queue size per flow when the round-trip is uniform.

verifies that the steady-state average queue size depends only on the mean round-trip delay.

The numerical examples presented here support the conclusions that (i) the oscillation in the queue size per flow decreases with increasing N, (ii) the average 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.

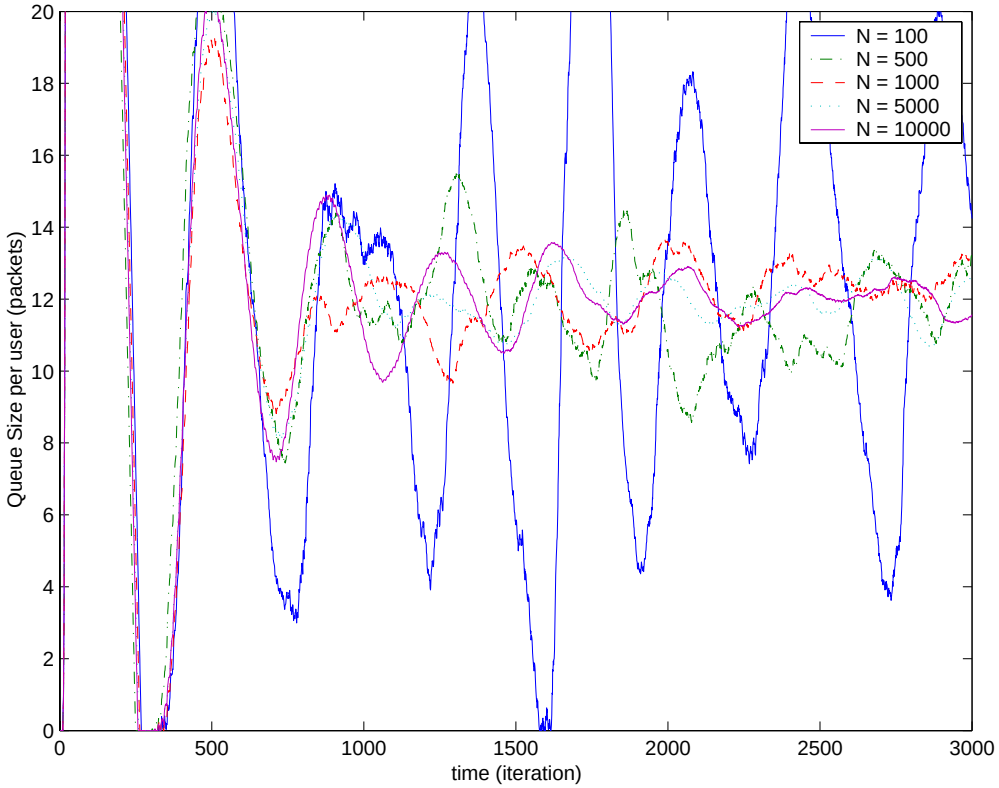

Figure 5.4: Example (ii) : Evolution of queue size per flow when the round-trip is Bernoulli.

NS-2 simulation results

In this section we verify our analysis by a more realistic event-driven NS 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 is configured as follows: $f^{(N)}(x) = f(N^{-1}x)$ (*i.e.,* a scaling similar to Assumption (AW1)) with $q_{min}^{(N)} = 2 \cdot N, q_{max}^{(N)} = 10 \cdot N$, and $p_{max} = 0.1$ 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 geometrically distributed workload that with a mean of 100 packets, and the interarrival times of the new workloads for each session 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 senders.

The simulation results are obtained with two types of 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 5.5. Next, the round-trip propagation delays of the sessions are randomly selected to be either 52 ms or 121.5 ms with equal probability (*i.e.,* i.i.d. Bernoulli rv). This delay distribution also has the mean of 87 ms but with much higher variance. Figure 5.6 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 5.6.

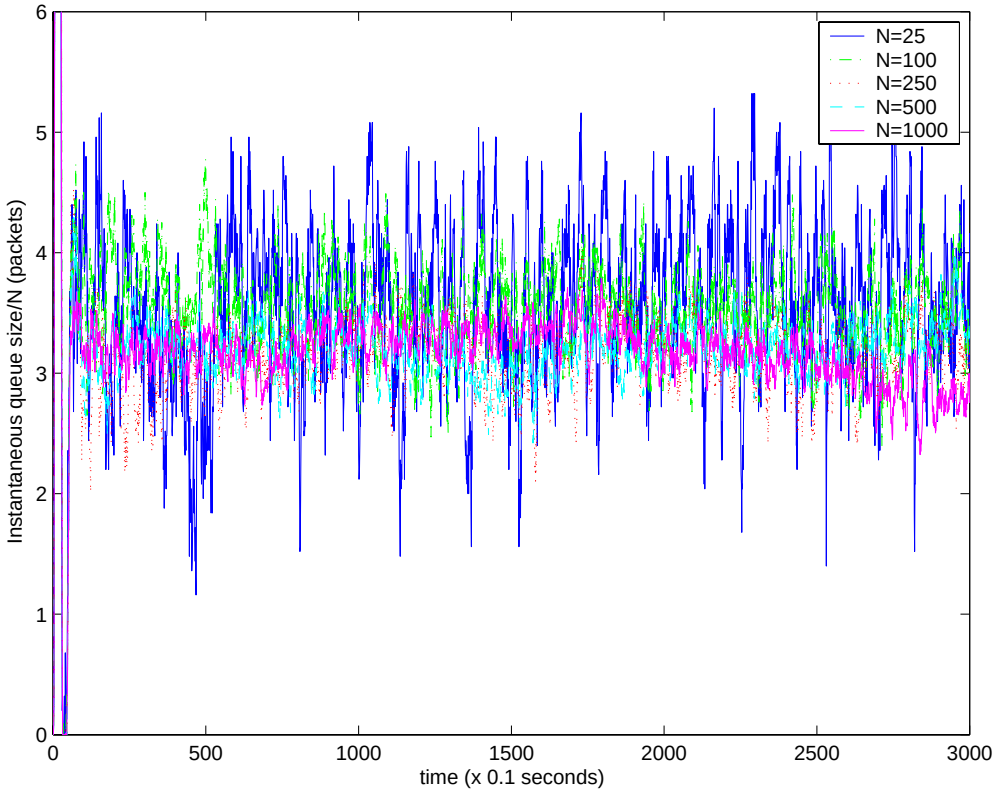

Figure 5.5: Queue dynamics of the NS-2 simulation where the round-trip distribution is uniform

5.7 A Proof of Lemma 1

First, the marking probability is fixed by (AW4b) 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, $\mathbf{P}\left[\min(W^\star, X^\star) = w | D^\star = d, \text{active}\right] = \mathbf{P}\left[\min(W^\star, X^\star) = w | \text{active}\right]$. 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 (AW4b)) and rv $\min(W^\star, X^\star)$ in between the transmission

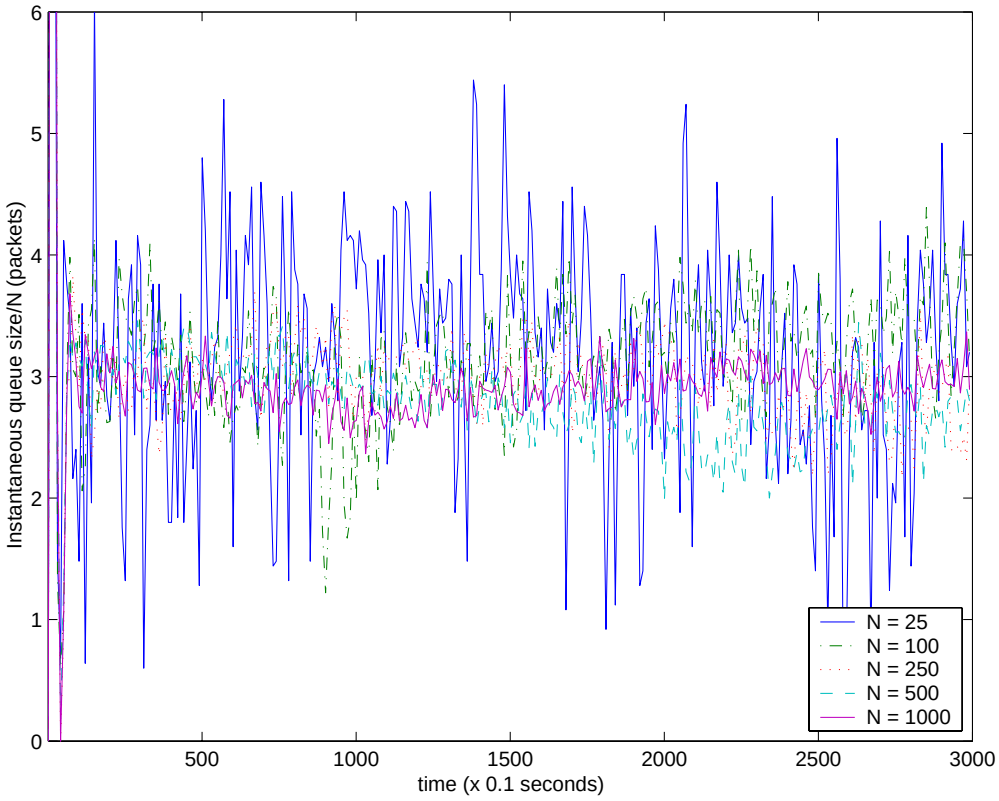

Figure 5.6: Queue dynamics of the NS-2 simulation where the round-trip distribution is Bernoulli

time has the same value as $\min(W^\star, X^\star)$ at the latest transmission time.

This also implies

$$\mathbf{P}\left[D^\star = d | \min(W^\star, X^\star) = w, \text{active}\right] = \mathbf{P}\left[D^\star = d | \text{active}\right].^{9}$$

Furthermore,

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

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

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

$\min(W^\star, X^\star)$. Finally,

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

and the desired result follows directly from the last equality.

Chapter 6

Weak Laws of Large Numbers

In this chapter, we present the proofs of the Weak Laws of Large Numbers for the window-based models in Chapters 4 and 5, *i.e.,* Theorems 2 and 4. The proofs of these theorems utilize similar ideas revolving around induction of convergence statements at each time $t = 1, 2, \ldots$. The recursion of the asymptotic queue size $q(t)$ is established as a byproduct of the proofs of the convergence statements.

Before presenting the proofs of Theorems 2 and 4, we first present in Section 6.1 a generic model of congestion-controlled flows and RED and then state a Weak Law of Large Numbers for this model (Theorem 6). Next, we show that the convergence statements in Theorems 2 and 4 are mere corollaries of Theorem 6 in Section 6.2 and 6.3, respectively. The proof of Theorem 6 is presented in Section 6.4. Finally, Section 6.5 establishes the distributional recursions in Theorems 2 and 4 as a byproduct of the proof of Theorem 6.

6.1 A Weak Law of Large Numbers of a Generic Model

In this section, we develop a generic model of congestion-controlled flows and a RED gateway. A Weak Law of Large Numbers for the model is then presented.

This result not only helps facilitate the presentation of the proofs of Theorems 2 and 4, but also yields insights on the crucial elements in the development of these limit theorems.

We denote the vector of state variables for session i in timeslot $[t, t+1)$ by $\mathbf{Y}_i^{(N)}(t)$. The rv $\mathbf{Y}_i^{(N)}(t)$ takes values in a discrete space $\mathcal{Y}$. Recall the recursion of the queue is given by (1.9), where $A_i^{(N)}(t)$ represents the amount of packets injected into the network in timeslot $[t, t+1)$ by session i and depends on the particular model. More specifically, the rv $A_i^{(N)}(t)$ is determined as a function of $\mathbf{Y}_i^{(N)}(t)$, *i.e.,* $A_i^{(N)}(t) = \Phi(\mathbf{Y}_i^{(N)}(t))$, for some bounded continuous function $\Phi : \mathcal{Y} \to \mathbb{R}$. Further, the recursion of the queue average used in the marking mechanism is given by (5.12). A special case when $\alpha = 1$ is equivalent to using the instantaneous queue size for marking mechanism, *e.g.,* the model in Theorem 2.

The marking mechanisms in Chapter 4 and 5 are similar and will be briefly summarized here. 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 $\hat{Q}^{(N)}(t)$ 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 $\{V_{i,j}(t+1),\ i, j = 1, \cdots;\ t = 0, 1, \cdots\}$. Given $t = 0, 1, \ldots$, the rvs $\{V_{i,j}(t+1),\ i, j = 1, \cdots\}$ are assumed to be independent of all the events happening prior to the beginning of timeslot $[t+1, t+2)$. For each $i = 1, \ldots, N$ and $j = 1, 2, \ldots$, we define the marking rvs $M_{i,j}^{(N)}(t+1)$ as in (5.13), 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 $M_{i,new}^{(N)}(t+1)$ given in (5.14).

With $\mathcal{F}_t = \sigma\left\{Q^{(N)}(0), \mathbf{Y}_i^{(N)}(0), V_{i,j}(s),\ 1 \le s \le t;\ i,j = 1,2,\ldots\right\}$, define

$$\begin{aligned} Z_i^{(N)}(t) &:= \mathbf{E}\left[M_{i,new}^{(N)}(t+1)|\mathcal{F}_t\right] \\ &= \left(1 - f^{(N)}(\hat{Q}^{(N)}(t))\right)^{A_i^{(N)}(t)}. \end{aligned} \tag{6.1}$$

In other words, $Z_i^{(N)}(t)$ is the conditional probability that, given $\mathcal{F}_t$, connection i will receive no marks during timeslot $[t, t+1)$.

Theorem 6 is discussed under the following assumptions.

(G1) There exists a continuous function $f : \mathbb{R}_+ \to [0,1]$ such that for each $N = 1, 2, \ldots$, Equation (1.1) holds.

(G2) For each $N = 1, 2, \ldots$, the dynamics (1.9) and (5.12) start with the conditions

$$Q^{(N)}(0) = \hat{Q}^{(N)}(t) = 0,$$

and

$$\mathbf{Y}_i^{(N)}(0) = \mathbf{y}, \quad i = 1, \ldots, N$$

for some $\mathbf{y}$ in $\mathcal{Y}$.

(G3) For any bounded mapping $g : \mathcal{Y} \to \mathbb{R}$, there exists a bounded and continuous mapping $F_g : [0,1] \times \mathcal{Y} \to \mathbb{R}$ such that

$$\mathbf{E}\left[g\left(\mathbf{Y}_i^{(N)}(t+1)\right)|\mathcal{F}_t\right] = F_g\left(Z_i^{(N)}(t), \mathbf{Y}_i^{(N)}(t)\right). \tag{6.2}$$

Assumption (G1) is identical to Assumption (AW1) in Theorems 2 and 4, while Assumption (G2) generalizes Assumption (AW2) and (A2Wb) in Theorems 2 and 4, respectively. It implies that for any $t = 0, 1, \ldots$ the collection of rvs $\mathbf{Y}_1^{(N)}(t), \ldots, \mathbf{Y}_N^{(N)}(t)$ are exchangeable. We will show in the discussion of Theorems 2 and 4 that Assumption (G3) indeed holds. Assumption (G3) states that given the events leading up to the beginning of timeslot $[t, t+1)$, the expected behavior of session i leading to the beginning of timeslot $[t+1, t+2)$ can also be determined using knowledge of the conditional marking probability $Z_i^{(N)}(t)$ and of $\mathbf{Y}_i^{(N)}(t)$. Note that (G3) immediately implies

$$\mathbf{E}\left[g\left(\mathbf{Y}_i^{(N)}(t+1)\right)\right] = \mathbf{E}\left[F_g\left(Z_i^{(N)}(t), \mathbf{Y}_i^{(N)}(t)\right)\right]. \tag{6.3}$$

Theorem 6. *Assume that (G1)-(G3) hold. Then, for each $N = 1, 2, \ldots$ and $t = 0, 1, \ldots$, there exists (non-random) constants $q(t)$, $\hat{q}(t)$ and a $\mathcal{Y}$-valued rv such that the following holds:*

(i) The following convergences take place:

$$\frac{Q^{(N)}(t)}{N} \quad \xrightarrow{P}_N \quad q(t) \tag{6.4}$$

$$\frac{\hat{Q}^{(N)}(t)}{N} \quad \xrightarrow{P}_N \quad \hat{q}(t) \tag{6.5}$$

and

$$\mathbf{Y}_1^{(N)}(t) \quad \Longrightarrow_N \quad \mathbf{Y}(t). \tag{6.6}$$

(ii) For any bounded function $g : \mathcal{Y} \to \mathbb{R}$,

$$\tfrac{1}{N}\textstyle\sum_{i=1}^N g\left(\mathbf{Y}_i^{(N)}(t)\right) \xrightarrow{P}_N \mathbf{E}\left[g\left(\mathbf{Y}(t)\right)\right]. \tag{6.7}$$

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

$$\lim_{N\to\infty} \mathbf{P}[\mathbf{Y}_i^{(N)}(t) = \mathbf{y}_i,\ i = 1, \cdots, I] \ = \ \prod_{i=1}^{I} \mathbf{P}\left[\mathbf{Y}(t) = \mathbf{y}_i\right] \qquad (6.8)$$

for any $\mathbf{y}_i \in \mathcal{Y}$, $i = 1, \ldots, I$.

In particular, if we take $g = \Phi$ in (6.7), we find

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} A_i^{(N)}(t) \quad \stackrel{P}{\to}_N \quad \mathbf{E}\left[A(t)\right]. \qquad (6.9)$$

From Theorem 6, we only need to show that the models in Theorems 2 and 4 both satisfy Assumption (G3), *i.e.,* identify the mapping F_g in (6.2), and subsequently establish the distributional recursions to finish the proofs of these theorems. The proofs of Theorems 2 and 4 will be presented following this approach in Section 6.2 and 6.3, respectively.

6.2 Proof of Theorem 2

According to the model in Chapter 4, the state variable $\mathbf{Y}_i^{(N)}(t)$ is $W_i^{(N)}(t)$ and the state space $\mathcal{Y}$ is the set $\{1, \ldots, W_{\max}\}$. The exponential weighted average parameter is $\alpha = 1$ and the average queue length $\hat{Q}^{(N)}(t)$ used for the marking mechanism coincides with the instantaneous queue length $Q^{(N)}(t)$ for all $t = 0, 1, \ldots$. The number of packets injected into the network $A_i^{(N)}(t)$ is given by $W_i^{(N)}(t)$. It still remains to show that Assumption (G3) in Theorem 6 and the distributional recursion (4.10)-(4.11) hold.

For some positive integer $p \geq 1$, consider an arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$: With g we associate the mappings $g^\star, g_\star : \mathbb{N} \to \mathbb{R}^p$ given by

$$g^\star(w) := g(\min\left(w + 1, W_{\max}\right)), \quad w \in \mathbb{N} \qquad (6.10)$$

and

$$g_\star(w) := g(\min\left(\lceil\frac{w}{2}\rceil, W_{\max}\right)), \quad w \in \mathbb{N}. \tag{6.11}$$

Now fix $i = 1, \ldots, N$ and $t = 0, 1, \ldots$. It follows from (4.5) that

$$\begin{aligned} & g(W_i^{(N)}(t+1)) \\ = \;& M_i^{(N)}(t+1) g^\star(W_i^{(N)}(t)) + (1 - M_i^{(N)}(t+1)) g_\star(W_i^{(N)}(t)) \end{aligned} \tag{6.12}$$

Let $\mathcal{F}_t$ denote the σ-field generated by the rvs

$$\{Q^{(N)}(0), W_i^{(N)}(0), V_i(s), V_{i,j}(s),\ i, j = 1, 2, \ldots;\ s = 1, \ldots, t\}.$$

The rvs $Q^{(N)}(t)$ and $W_i^{(N)}(t),\ i = 1, \ldots, N$, being all $\mathcal{F}_t$-measurable, it holds under the enforced independence assumptions that

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

so that

$$\mathbf{E}\left[M_i^{(N)}(t+1)|\mathcal{F}_t\right] = Z_i^{(N)}(t) \tag{6.13}$$

by conditional independence, where we have set

$$Z_i^{(N)}(t) = \left(1 - f^{(N)}(Q^{(N)}(t))\right)^{W_i^{(N)}(t)}. \tag{6.14}$$

It is now plain that

$$M_i^{(N)}(t+1) =_{st} \mathbf{1}\left[V_i(t+1) \leq Z_i^{(N)}(t)\right]. \tag{6.15}$$

It readily follows from (6.12) that

$$\mathbf{E}\left[g(W_i^{(N)}(t+1))|\mathcal{F}_t\right] = F_g(Z_i^{(N)}(t), W_i^{(N)}(t)) \tag{6.16}$$

where the mapping $F_g : [0, 1] \times \mathbb{N} \to \mathbb{R}^p$ is associated with g through

$$F_g(z, w) \;=\; z g^\star(w) + (1 - z) g_\star(w), \quad z \in [0, 1],\ w \in \mathbb{N}. \tag{6.17}$$

Thus, we can take the mapping specified by (6.17) to be the mapping F_g which ensures (6.2) and Assumption (G3) thus holds. The convergence statements in Theorem 2 is now a direct corollary of Theorem 6. It remains only to establish the distributional recursions (4.9)-(4.11). This will be accomplished in Section 6.5 as it is a byproduct of the proof of Theorem 6.

6.3 Proof of Theorem 4

The state variable $\mathbf{Y}_i^{(N)}(t)$ in the model of Chapter 5 is given in (5.16) with the state space $\mathcal{Y}$ given by (5.17). In the model, the exponential weighted average parameter lies in the range $(0, 1]$ and the average queue length $\hat{Q}^{(N)}(t)$ is used for the marking mechanism instead of the instantaneous queue length $Q^{(N)}(t)$ for all $t = 0, 1, \ldots$. Finally, we have $A_i^{(N)}(t)$ given by (5.5).

Fix $i = 1, \ldots, N$ and consider an arbitrary *bounded* mapping $g : \mathbb{Z}_+^6 \to \mathbb{R}$:

Through a careful case analysis, it follows that

$$
\begin{aligned}
& g(\mathbf{Y}_i^{(N)}(t+1)) \qquad (6.18)\\
&= \mathbf{1}\left[X_i^{(N)}(t)=0\right] \\
&\quad \times g\left(0, \mathbf{1}\left[U_i(t+1)<P_{ar}\right]F_i(t+1), 1, \mathbf{1}\left[U_i(t+1)<P_{ar}\right]D_{i,new}(t+1), 0, 1\right) \\
&+ \mathbf{1}\left[X_i^{(N)}(t)>A_i^{(N)}(t)>0\right] \\
&\quad \times g(W_i^{(N)}(t), X_i^{(N)}(t)-A_i^{(N)}(t), M_{i,new}^{(N)}(t+1)S_i^{(N)}(t), D_i^{(N)}(t), 1, M_{i,new}^{(N)}(t+1)) \\
&+ \mathbf{1}\left[0<X_i^{(N)}(t)\le A_i^{(N)}(t)\right] g(0,0,1,D_i^{(N)}(t),1,M_{i,new}^{(N)}(t+1)) \\
&+ \mathbf{1}\left[X_i^{(N)}(t)>0, \beta_i^{(N)}(t)=D_i^{(N)}(t)-1\right] \\
&\quad \times g(W_{i,new}^{(N)}(t+1), X_i^{(N)}(t), S_i^{(N)}(t), D_i^{(N)}(t), D_i^{(N)}(t), M_i^{(N)}(t)) \\
&+ \mathbf{1}\left[X_i^{(N)}(t)>0, \beta_i^{(N)}(t)<D_i^{(N)}(t)-1\right] \\
&\quad \times g(W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), D_i^{(N)}(t), \beta_i^{(N)}(t)+1, M_i(t))
\end{aligned}
$$

where

$$
\begin{aligned}
& g(W_{i,new}^{(N)}(t+1), X_i^{(N)}(t), S_i^{(N)}(t), D_i^{(N)}(t), D_i^{(N)}(t), M_i^{(N)}(t)) \\
= \;& M_i^{(N)}(t)S_i^{(N)}(t)F_g^1\left(W_i^{(N)}(t), X_i^{(N)}(t), D_i^{(N)}(t)\right) \\
+ \;& M_i^{(N)}(t)(1-S_i^{(N)}(t))F_g^2\left(W_i^{(N)}(t), X_i^{(N)}(t), D_i^{(N)}(t)\right) \\
+ \;& (1-M_i^{(N)}(t))F_g^3(W_i^{(N)}(t), X_i^{(N)}(t), D_i^{(N)}(t)). \qquad (6.19)
\end{aligned}
$$

The mappings F_g^1, F_g^2 and $F_g^3 : \mathbb{Z}_+^3 \to \mathbb{R}$ are associated with g and defined by:

$$
\begin{aligned}
F_g^1(w,x,d) &= g\left(\min\left(2w \vee 1, W_{\max}\right), x, 1, d, d, 1\right), \\
F_g^2(w,x,d) &= g\left(\min\left(w+1, W_{\max}\right), x, 0, d, d, 1\right), \\
F_g^3(w,x,d) &= g\left(\lceil \frac{w}{2} \rceil, x, 0, d, d, 0\right). \qquad (6.20)
\end{aligned}
$$

Let $\mathcal{F}_t$ denote the σ-field generated by the rvs $\{Q^{(N)}(0), \mathbf{Y}_i^{(N)}(0), U_i(s), F_i(s), D_{i,new}(s), V_i(s), V_{i,j}(s)\ , i,j = 1,2,\ldots ; s = 1,\ldots,t\ \}$. So the rvs $Q^{(N)}(t)$ and

$\mathbf{Y}_i^{(N)}(t)$ $(i = 1, \ldots, N)$ are $\mathcal{F}_t$-measurable. Under the enforced independence assumptions, it holds that

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

Therefore,

$$\mathbf{E}\left[\prod_{i=1}^{A(t)} M_{i,j}^{(N)}(t+1)|\mathcal{F}_t\right] = Z_i^{(N)}(t) \tag{6.21}$$

by conditional independence, where we have set

$$Z_i^{(N)}(t) = \left(1 - f^{(N)}(\hat{Q}^{(N)}(t))\right)^{A_i^{(N)}(t)}. \tag{6.22}$$

It is now clear that

$$\prod_{i=1}^{A(t)} M_{i,j}^{(N)}(t+1) =_{st} \mathbf{1}\left[V_i(t+1) \leq Z_i^{(N)}(t)\right]. \tag{6.23}$$

Finally, it readily follows from (6.18) that

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

$$= F_g(Z_i^{(N)}(t), W_i^{(N)}(t), X_i^{(N)}(t), S_i^{(N)}(t), D_i^{(N)}(t), \beta_i^{(N)}(t), M_i^{(N)}(t)) \quad (6.24)$$

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

$$\begin{aligned}
&F_g(z,w,x,s,d,b,m) \qquad (6.25)\\
&= \mathbf{1}[x=0]\\
&\quad \times \mathbf{E}\left[g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] D_{i,new}(t+1), 0, 1\right)\right]\\
&+ \mathbf{1}[x > \min(w,x)\, \mathbf{1}[b \geq d > 0]]\\
&\quad \times [zg(w, x - \min(w,x)\, \mathbf{1}[b \geq d > 0], s, d, 1, 1)\\
&\quad +(1-z)g(w, x - \min(w,x)\, \mathbf{1}[b \geq d > 0], 0, d, 1, 0)]\\
&+ \mathbf{1}[0 < x \leq \min(w,x)\, \mathbf{1}[b \geq d > 0]]\,[zg(0,0,1,d,1,1) + (1-z)g(0,0,1,d,1,0)]\\
&+ \mathbf{1}[x > 0, b = d-1]\, g_{new}(w,x,s,d,m)\\
&+ \mathbf{1}[x > 0, b < d-1]\, g(w,x,s,d,b+1,m),
\end{aligned}$$

with

$$g_{new}(w,x,s,d) \;=\; msF_g^1(w,x,d) + m(1-s)F_g^2(w,x,d) + (1-m)F_g^3(w,x,d).$$

We note that

$$\mathbf{E}\left[g\left(0, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] F_i(t+1), 1, \mathbf{1}\left[U_i(t+1) < P_{ar}\right] D_{i,new}(t+1), 0, 1\right)\right]$$

always exists and is finite by the boundedness of the mapping g. Furthermore, the mapping F_g is continuous with respect to the product topology on the set $\mathcal{Y}$. Thus, the mapping F_g specified by (6.25) ensures (6.2) and Assumption (G3) thus holds. The convergence statements in Theorem 4 is again a direct corollary of Theorem 6. The distributional recursions for Theorem 4 will be established in Section 6.5.

6.4 A Proof of Theorem 6

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**] Equations (6.4) and (6.5) hold for some non-random constants $q(t)$ and $\hat{q}(t)$.

[**B:t**] Equation (6.6) holds for a $\mathcal{Y}$-valued rv $\mathbf{Y}(t)$.

[**C:t**] For any integer $I = 1, 2, \ldots$, the rv $\{\mathbf{Y}_i^{(N)}(t),\ i = 1, \ldots, I\}$ become asymptotically independent with large N as described by (6.8), where $\mathbf{Y}(t)$ are the rvs occurring in [**B:t**].

[**D:t**] For any bounded mapping $g : \mathcal{Y} \to \mathbb{R}$, the convergence (6.7) holds with $\mathbf{Y}(t)$ being the rv occurring in [**B:t**].

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 establish Theorem 6.

Lemma 2. *Under (G1) and (G3), if* [**A:t**] *and* [**B:t**] *hold for some* $t = 0, 1, \ldots$, *then* [**B:t+1**] *holds.*

Proof. Together the convergence [**A:t**] and [**B:t**] imply [29, Thm. 5.28, p. 150] the joint convergence

$$(N^{-1}\hat{Q}^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \implies_N (\hat{q}(t), \mathbf{Y}(t)). \tag{6.26}$$

Next the continuity of the mappings f implies that of $(x, \mathbf{y}) \to (1 - f(x))^{\Phi(\mathbf{y})}$ on $\mathbb{R}_+ \times \mathcal{Y}$, so that

$$(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \Longrightarrow_N (Z(t), \mathbf{Y}(t)) \tag{6.27}$$

by the Continuous Mapping Theorem [29, Thm. 5.29, p. 150] with

$$Z(t) = (1 - f(\hat{q}(t)))^{\Phi(\mathbf{Y}(t))} \ .$$

Consider (6.3) for any *bounded* arbitrary mapping $g : \mathcal{Y} \to \mathbb{R}$, and recall that the mapping F_g defined by (6.2) is continuous on $[0,1] \times \mathcal{Y}$. Consequently, the Continuous Mapping Theorem can again be invoked to yield

$$F_g(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t)) \implies_N F_g(Z(t), \mathbf{Y}(t)), \tag{6.28}$$

whence

$$\lim_{N\to\infty} \mathbf{E}\left[F_g(Z_1^{(N)}(t), \mathbf{Y}_1^{(N)}(t))\right] = \mathbf{E}\left[F_g(Z(t), \mathbf{Y}(t))\right] \tag{6.29}$$

by the Bounded Convergence Theorem [29, Thm. 4.16, p. 108]. Combining (6.3) and (6.29) we get

$$\lim_{N\to\infty} \mathbf{E}\left[g(\mathbf{Y}_1^{(N)}(t))\right] = \mathbf{E}\left[F_g(Z(t), \mathbf{Y}(t))\right] \tag{6.30}$$

and the bounded mapping g being arbitrary, it follows immediately that

$$\mathbf{Y}_1^{(N)}(t+1) \implies_N \mathbf{Y}(t+1)$$

for some $\mathcal{Y}$-valued rv $\mathbf{Y}(t+1)$ characterized by

$$\mathbf{E}\left[g\left(\mathbf{Y}(t+1)\right)\right] = \mathbf{E}\left[F_g\left(Z(t), \mathbf{Y}(t)\right)\right]. \tag{6.31}$$

□

Lemma 3. *Under (G1), (G3), 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**] (in particular (6.9)), we conclude that

$$\frac{Q^{(N)}(t)}{N} - C + \frac{1}{N}\sum_{i=1}^{N} A_i^{(N)}(t) \quad \stackrel{P}{\to}_N \quad q(t) - C + \mathbf{E}\left[A(t)\right] \tag{6.32}$$

and the desired result is a simple consequence of the continuity of the function $x \to x^+$ 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 $N = 1, 2, \ldots$. Therefore,

$$q(t+1) = [q(t) - C + \mathbf{E}\left[A(t)\right]]^+. \tag{6.33}$$

Also $\frac{\hat{Q}^{(N)}(t)}{N} \stackrel{P}{\to}_N \hat{q}(t)$ from [**A:t**], then it is simple to see that

$$\begin{aligned} \frac{\hat{Q}^{(N)}(t+1)}{N} &= (1-\alpha)\frac{\hat{Q}^{(N)}(t)}{N} + \alpha\frac{Q^{(N)}(t+1)}{N} \\ &\stackrel{P}{\to}_N (1-\alpha)\hat{q}(t) + \alpha q(t+1) \\ &= \hat{q}(t+1). \end{aligned} \tag{6.34}$$

□

Lemma 4. *Under (G1)–(G3), 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 from (6.2) that for a fixed N, the rvs $\mathbf{Y}_i^{(N)}(t+1), i = 1, \ldots, N$ are coupled only through the marking probability which depends only on $\hat{Q}^{(N)}(t)$. Fix a positive integer I. The rvs $\mathbf{Y}_1^{(N)}(t+1), \ldots, \mathbf{Y}_I^{(N)}(t+1)$ are mutually independent given $\mathcal{F}_t$. Consequently, for arbitrary bounded mappings $g_1, \ldots, g_I : \mathcal{Y} \to \mathbb{R}$, we get

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

with the help of (G3).

Now it follows from (6.8) in **[C:t]** that the joint convergence

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

holds where limiting rvs $\mathbf{Y}_1(t) \ldots, \mathbf{Y}_I(t)$ 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 (6.28) also lead to

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

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,

$$\begin{aligned}
\lim_{N\to\infty} \mathbf{E}\left[\prod_{i=1}^{I} g_i(\mathbf{Y}_i^{(N)}(t+1))\right] &= \lim_{N\to\infty} \mathbf{E}\left[\prod_{i=1}^{I} F_{g_i}(Z_i^{(N)}(t), \mathbf{Y}_i^{(N)}(t))\right] \\
&= \mathbf{E}\left[\prod_{i=1}^{I} F_{g_i}(Z_i(t), \mathbf{Y}_i(t))\right] \\
&= \prod_{i=1}^{I} \mathbf{E}\left[F_{g_i}(Z_i(t), \mathbf{Y}_i(t))\right] \\
&= \prod_{i=1}^{I} \mathbf{E}\left[g_i(\mathbf{Y}_i(t))\right] \qquad (6.35)
\end{aligned}$$

where the last equality made use of the relation (6.31). The desired result **[C:t+1]** now follows from (6.35) given that the mappings $g_1, \ldots, g_I$ are arbitrary. □

Lemma 5. *Under (G1)–(G3), if* **[A:t]**, **[B:t]** *and* **[C:t]** *hold for some* $t = 0, 1, \ldots$, *then* **[D:t]** *holds.*

Proof. Pick a bounded mapping $g : \mathcal{Y} \to \mathbb{R}$. Under (G2) the rvs $\mathbf{Y}_i^{(N)}(t)$, $i = 1, \ldots, N$, are exchangeable, so that

$$\begin{aligned} & \operatorname{var}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t))\right] \\ = \; & N^{-2}\sum_{i=1}^{N} \operatorname{var}[g(\mathbf{Y}_i^{(N)}(t))] \\ + \; & N^{-2}\sum_{i,j=1, i\neq j}^{N} \operatorname{cov}[g(\mathbf{Y}_i^{(N)}(t)), g(\mathbf{Y}_j^{(N)}(t))] \\ = \; & N^{-1}\operatorname{var}[g(\mathbf{Y}_1^{(N)}(t))] + \frac{N-1}{N}\operatorname{cov}[g(\mathbf{Y}_1^{(N)}(t)), g(\mathbf{Y}_2^{(N)}(t))]. \end{aligned} \tag{6.36}$$

Now let N go to infinity in (6.36): The validity of **[C:t]** and the Bounded Convergence Theorem already imply

$$\lim_{N\to\infty} \operatorname{cov}[g(\mathbf{Y}_1^{(N)}(t)), g(\mathbf{Y}_2^{(N)}(t))] \;=\; \operatorname{cov}[g(\mathbf{Y}_1(t)), g(\mathbf{Y}_2(t))] = 0 \tag{6.37}$$

by asymptotic independence. On the other hand,

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

since g is bounded.

Combining these observations we readily see that

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

whence, by Chebyshev's inequality,

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t)) - \mathbf{E}\left[\tfrac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t))\right] \quad \xrightarrow{P}_N \quad 0. \tag{6.38}$$

This last convergence is equivalent to

$$\tfrac{1}{N}\textstyle\sum_{i=1}^{N} g(\mathbf{Y}_i^{(N)}(t)) - \mathbf{E}\left[g(\mathbf{Y}_1^{(N)}(t))\right] \xrightarrow{P}_N 0$$

by exchangeability, and the desired convergence result (6.7) is now immediate once we remark under **[B:t]** that $\lim_{N\to\infty} \mathbf{E}\left[g(\mathbf{Y}_1^{(N)}(t))\right] = \mathbf{E}\left[g(\mathbf{Y}(t))\right]$. □

We now conclude with a proof of Theorem 6: We first note that under (G1)-(G3) 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 6 is now established.

6.5 A Note on the Distributional Recursions

The proof of Lemma 3 also shows that

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

and

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

with non-random $q(t+1)$ and $\hat{q}(t+1)$ determined by (6.33) and (6.34), respectively. Therefore, the asymptotic queue recursions (4.9) in Chapter 4 and (5.22)-(5.23) in Chapter 5 hold.

A moment of reflection and a comparison to the decomposition in (6.16)-(6.17) and the analysis leading up to (6.31) in Lemma 2 will convince the reader that the distributional recursions (4.10)-(4.11) in Theorem 2 hold. Similarly, by using the decomposition (6.24)-(6.25), we establish the distributional recursion (5.24) in Theorem 4.

Chapter 7

A Proof of Central Limit Theorem

In this chapter, we present the proof of the Central Limit Theorem for the window-based model in Chapter 4, *i.e.,* Theorem 3. As in the proof of Theorem 6, we shall proceed by induction on t with the help of a series of technical facts. Again the discussion is facilitated by introducing for each $t = 0, 1, \ldots$, a number of auxiliary convergence statements.

Throughout the discussion, for each $t = 0, 1, \ldots$, we shall find it useful to write

$$Z(t) = (1 - f(q(t)))^{W(t)} = \gamma(t)^{W(t)} \tag{7.1}$$

with

$$\gamma(t) = 1 - f(q(t)). \tag{7.2}$$

Also we note a simple but useful consequence of Claim (i) of Theorem 2, namely that

$$f(\frac{Q^{(N)}(t)}{N}) \xrightarrow{P}_N f(q(t)) \tag{7.3}$$

by the continuity of f.

With the aim to simplify the presentation, for arbitrary integer p and

arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$, we define

$$L_g^{(N)}(t) := \frac{1}{N}\left(\sum_{i=1}^{N} g(W_i^{(N)}(t)) - \mathbf{E}\left[g(W(t))\right]\right), \quad N = 1, 2, \ldots$$

for each $t = 0, 1, \ldots$. Note that $\bar{L}^{(N)}(t)$ corresponds to the choice $g(w) = w$.

The discussion is facilitated by introducing for each $t = 0, 1, \ldots$, a number of auxiliary convergence statements, namely [**E:t**] and [**F:t**] where

[**E:t**] For arbitrary integer p and arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$, there exists an $\mathbb{R}^{p+1}$-valued rv $(L_0(t), L_g(t))$ such that the joint convergence

$$\sqrt{N}(L_0^{(N)}(t), L_g^{(N)}(t)) \Longrightarrow_N (L_0(t), L_g(t)) \tag{7.4}$$

takes place;

[**F:t**] For arbitrary integer p and arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$, there exists an $\mathbb{R}^{p+1}$-valued rv $(L_0(t+1), L_g(t))$ such that the joint convergence

$$\sqrt{N}(L_0^{(N)}(t+1), L_g^{(N)}(t)) \Longrightarrow_N (L_0(t+1), L_g(t)) \tag{7.5}$$

takes place with the rv $L_0(t+1)$ related to $L_0(t)$ through the distributional recurrence (4.26).

The convergence statements propagate in time as the discussion now shows:

Proposition 2. *Under (AW1b)–(AW2), if* [**E:t**] *holds for some* $t = 0, 1, \ldots$, *then* [**F:t**] *holds with* $\mathbb{R}$*-valued rv* $L_0(t+1)$ *satisfying the distributional recurrence (4.26).*

Proposition 3. *Under (AW1b)–(AW2), if* [**E:t**] *holds for some* $t = 0, 1, \ldots$, *then* [**E:t+1**] *also holds.*

In the process of establishing Proposition 3, we will obtain the following fact: For arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$, the limiting rv $L_g(t+1)$ has the following form

$$L_g(t+1) =_{st} L_{\hat{g}_t}(t) + f'(q(t))R_g(t)L_0(t) + H_{\hat{g}}(t+1) \tag{7.6}$$

where $R_g(t)$ is introduced at (7.26) and the $\mathbb{R}^p$-valued rv $H_{\hat{g}}(t+1)$ is a zero-mean Gaussian rv with covariance matrix

$$\Sigma_{\hat{g}}(t+1) := \mathbf{E}\left[\hat{g}(W(t))\hat{g}(W(t))'Z(t)(1-Z(t))\right] \tag{7.7}$$

and this Gaussian rv is independent of the rvs $L_{\hat{g}_t}(t)$ and $L_0(t)$. The mappings $\hat{g}$ and $\hat{g}_t$ are defined at (7.15) and (7.17), respectively.

We complete the proof of Theorem 3 by an easy induction argument: For $t = 0$, [**E:t**] trivially holds since for each $N = 1, 2, \ldots$, we have $W_i^{(N)}(0) = W$ for all $i = 1, \ldots, N$ and $W(0) = W$. It is now plain from Proposition 2 and Proposition 3 that [**E:t**] and [**F:t**] both hold for all $t = 0, 1, \ldots$, and Theorem 3 is established.

7.1 A Proof of Proposition 2

Fix $t = 0, 1, \ldots$ and $N = 1, 2, \ldots$. We begin by noting that under [**E:t**], the mapping g being *arbitrary*, it is the case that there exists an $\mathbb{R}^{p+2}$-valued rv $(L_0(t), \bar{L}(t), L_g(t))$ such that the joint convergence

$$\sqrt{N}(L_0^{(N)}(t), \bar{L}^{(N)}(t), L_g^{(N)}(t)) \Longrightarrow_N (L_0(t), \bar{L}(t), L_g(t)) \tag{7.8}$$

takes place. Indeed it suffices to use [**E:t**] with the mapping $w \to (w, g(w))$.

As we seek to identify $L_0(t+1)$, we rewrite the limiting recursion (4.9) in the

form

$$q(t+1) = (q(t) - C + \mathbf{E}\left[W(t)\right])^+ = (-K(t))^+ \tag{7.9}$$

with $K(t)$ given by (4.27). Combining this observation with the queue dynamics (4.2), we get

$$\begin{aligned} L_0^{(N)}(t+1) &= \left(\frac{Q^{(N)}(t)}{N} - C + \frac{1}{N}\sum_{i=1}^{N} W_i^{(N)}(t) \right)^+ - (-K(t))^+ \\ &= \max\left(L_0^{(N)}(t) + \frac{1}{N}\sum_{i=1}^{N} W_i^{(N)}(t) - \mathbf{E}\left[W(t)\right], K(t) \right) - K(t)^+ \\ &= \max\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t), K(t) \right) - K(t)^+ \end{aligned} \tag{7.10}$$

so that

$$\begin{aligned} & \sqrt{N} L_0^{(N)}(t+1) \\ = \; & \max\left(\sqrt{N}\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t) \right), \sqrt{N} K(t) \right) - \sqrt{N} K(t)^+. \end{aligned} \tag{7.11}$$

Invoking the Continuous Mapping Theorem, we conclude from (7.8) that

$$\sqrt{N}\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t) \right) \Longrightarrow_N L_0(t) + \bar{L}(t). \tag{7.12}$$

Three cases emerge depending on the sign of $K(t)$. If $K(t) = 0$, then (7.11) reduces to

$$\sqrt{N} L_0^{(N)}(t+1) = \left(\sqrt{N}\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t) \right) \right)^+ \tag{7.13}$$

again by the Continuous Mapping Theorem and the convergence (7.12) yields

$$\sqrt{N} L_0^{(N)}(t+1) \Longrightarrow_N \left(L_0(t) + \bar{L}(t) \right)^+.$$

If $K(t) < 0$, then (7.11) reduces to

$$\sqrt{N} L_0^{(N)}(t+1) = \max\left(\sqrt{N}\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t) \right), -\sqrt{N}|K(t)| \right)$$

and the convergence (7.12) yields

$$\sqrt{N}L_0^{(N)}(t+1) \Longrightarrow_N L_0(t) + \bar{L}(t)$$

since $|K(t)| > 0$ guarantees $\lim_{N\to\infty} \sqrt{N}|K(t)| = \infty$.

Finally, if $K(t) > 0$, then (7.11) reduces to

$$\sqrt{N}L_0^{(N)}(t+1) = \max\left(\sqrt{N}\left(L_0^{(N)}(t) + \bar{L}^{(N)}(t)\right) - \sqrt{N}K(t), 0\right)$$

and the convergence (7.12) yields $\sqrt{N}L_0^{(N)}(t+1) \Longrightarrow_N 0$ since $\lim_{N\to\infty} \sqrt{N}K(t) = \infty$. This completes the proof of Proposition 2. □

7.2 A Key Decomposition

To establish Proposition 3, we start with an arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$ for some positive integer p, and define the mappings $g^\star, g_\star : \mathbb{N} \to \mathbb{R}^p$ by (6.10) and (6.11), respectively.

Fix $N = 1, 2, \ldots$, $i = 1, \ldots, N$ and $t = 0, 1, \ldots$. Making use of (4.5) and (4.10), we get

$$\begin{aligned}
& g(W_i^{(N)}(t+1)) - \mathbf{E}\left[g(W(t+1))\right] \\
= \; & M_i^{(N)}(t+1)g^\star(W_i^{(N)}(t)) + (1 - M_i^{(N)}(t+1))g_\star(W_i^{(N)}(t)) \\
& -\mathbf{E}\left[M(t+1)g^\star(W(t)) + (1 - M(t+1))g_\star(W(t))\right] \\
= \; & g_\star(W_i^{(N)}(t)) - \mathbf{E}\left[g_\star(W(t))\right] \\
& + \left(g^\star(W_i^{(N)}(t)) - g_\star(W_i^{(N)}(t))\right) M_i^{(N)}(t+1) \\
& -\mathbf{E}\left[\left(g^\star(W(t)) - g_\star(W(t))\right) M(t+1)\right]. \qquad (7.14)
\end{aligned}$$

Therefore, introducing the mapping $\hat{g} : \mathbb{N} \to \mathbb{R}^p$ defined by

$$\hat{g}(w) = g^\star(w) - g_\star(w), \quad w \in \mathbb{N} \tag{7.15}$$

we obtain the decomposition

$$\begin{aligned} & g(W_i^{(N)}(t+1)) - \mathbf{E}\left[g(W(t+1))\right] \\ = \; & g_\star(W_i^{(N)}(t)) - \mathbf{E}\left[g_\star(W(t))\right] \\ & + \hat{g}(W_i^{(N)}(t)) \left(M_i^{(N)}(t+1) - Z_i^{(N)}(t)\right) \\ & + \hat{g}(W_i^{(N)}(t)) \left(Z_i^{(N)}(t) - \gamma(t)^{W_i^{(N)}(t)}\right) \\ & + \hat{g}(W_i^{(N)}(t))\gamma(t)^{W_i^{(N)}(t)} - \mathbf{E}\left[\hat{g}(W(t))Z(t)\right]. \end{aligned} \tag{7.16}$$

Finally, defining the mapping $\hat{g}_t : \mathbb{N} \to \mathbb{R}^p$ by

$$\hat{g}_t(w) := g_\star(w) + \hat{g}(w) \cdot \gamma(t)^w, \quad w \in \mathbb{N} \tag{7.17}$$

we find

$$\begin{aligned} & g(W_i^{(N)}(t+1)) - \mathbf{E}\left[g(W(t+1))\right] \\ = \; & \hat{g}_t(W_i^{(N)}(t)) - \mathbf{E}\left[\hat{g}_t(W(t))\right] \\ & + \hat{g}(W_i^{(N)}(t)) \left(M_i^{(N)}(t+1) - Z_i^{(N)}(t)\right) \\ & + \hat{g}(W_i^{(N)}(t)) \left(Z_i^{(N)}(t) - \gamma(t)^{W_i^{(N)}(t)}\right). \end{aligned} \tag{7.18}$$

This decomposition forms the basis for the subsequent analysis. The subsequent sections discuss the needed asymptotics for each of the three terms of (7.18).

7.3 The First Piece of the Puzzle

To study the contribution of the second term of (7.18), we proceed as follows; For any mapping $g : \mathbb{N} \to \mathbb{R}^p$, set

$$H_g^{(N)}(t+1) := \frac{1}{N} \sum_{i=1}^{N} g(W_i^{(N)}(t)) \left(M_i^{(N)}(t+1) - Z_i^{(N)}(t) \right), \quad N = 1, 2, \ldots$$

for each $t = 0, 1, \ldots$. We begin with the case $p = 1$.

Proposition 4. *Assume (AW1b)-(AW2) to hold and consider an arbitrary mapping $g : \mathbb{N} \to \mathbb{R}$. Then, for each $t = 0, 1, \ldots$, it holds that*

$$\mathbf{E}\left[\exp\left(j\theta\sqrt{N} H_g^{(N)}(t+1) \right) | \mathcal{F}_t \right] \xrightarrow{P}_N e^{-\frac{\theta^2}{2}\sigma_g(t+1)}, \quad \theta \in \mathbb{R} \tag{7.19}$$

with

$$\sigma_g(t+1) := \mathbf{E}\left[g(W(t))^2 Z(t)(1 - Z(t)) \right]. \tag{7.20}$$

The proof of Proposition 4 is given in Section 7.8. By using the standard Cramer-Wold device [6, Thm. 7.7, p. 49] we obtain the following analog in higher dimensions:

Corollary 1. *Assume (AW1b)-(AW2) to hold and consider an arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$. Then, for each $t = 0, 1, \ldots$, it holds that*

$$\mathbf{E}\left[\exp\left(j\theta'\sqrt{N} H_g^{(N)}(t+1) \right) | \mathcal{F}_t \right] \xrightarrow{P}_N e^{-\frac{1}{2}\theta'\Sigma_g(t+1)\theta}, \quad \theta \in \mathbb{R}^p \tag{7.21}$$

with

$$\Sigma_g(t+1) := \mathbf{E}\left[g(W(t)) g(W(t))' Z(t)(1 - Z(t)) \right]. \tag{7.22}$$

We conclude with the following crucial by-products: For some $t = 0, 1, \ldots,$ consider the situation where a sequence of $\mathbb{R}^q$-valued rvs $\{\Lambda^{(N)}(t),\ N = 1, 2, \ldots\}$ weakly converges, say

$$\Lambda^{(N)}(t) \Longrightarrow_N \Lambda(t) \tag{7.23}$$

for some limiting $\mathbb{R}^q$-valued rv $\Lambda(t)$. If for each $N = 1, 2, \ldots$, the rv $\Lambda^{(N)}(t)$ is $\mathcal{F}_t$-measurable, then Corollary 1 readily implies [6, Thm. 3.2, p. 21] that

$$(\sqrt{N} H_g^{(N)}(t+1), \Lambda^{(N)}(t)) \Longrightarrow_N (H_g(t+1), \Lambda(t)) \tag{7.24}$$

for some zero-mean Gaussian rv $H_g(t+1)$ with covariance matrix $\Sigma_g(t+1)$. The rv $H_g(t+1)$ is taken to be *independent* of the rv $\Lambda(t)$.

7.4 The Delta Method

We begin by recalling a well-known byproduct of the Central Limit Theorem, known as the *Delta Method* [29, p. 214]. Its statement is given as Proposition 1 in Chapter 4. The proof is provided in Section 7.6 as it is crucial for establishing the joint convergence statement at the end of this section.

We are now in a position to handle the contributions of the last term of the decomposition (7.18). For any mapping $g : \mathbb{N} \to \mathbb{R}^p$, set

$$Z_g^{(N)}(t) := \frac{1}{N} \sum_{i=1}^{N} g(W_i^{(N)}(t)) \left(\gamma^{(N)}(t)^{W_i^{(N)}(t)} - \gamma(t)^{W_i^{(N)}(t)} \right), \quad N = 1, 2, \ldots$$

for each $t = 0, 1, \ldots$. The next result makes use of the Delta Method and its proof is available in Section 7.9.

Proposition 5. *Assume (AW1b)-(AW2) to hold and consider an arbitrary mapping $g : \mathbb{N} \to \mathbb{R}^p$. If for some $t = 0, 1, \ldots$, the convergence (4.30) holds with some*

rv $L_0(t)$*, then it holds that*

$$\sqrt{N}Z_g^{(N)}(t) \Longrightarrow_N f'(q(t))R_g(t)L_0(t) \tag{7.25}$$

with

$$R_g(t) := \mathbf{E}\left[W(t)\left(1 - f(q(t))\right)^{W(t)-1} g(W(t))\right]. \tag{7.26}$$

A careful inspection of the proof of Proposition 5 in Section 7.9 reveals that a somewhat stronger statement is true: For some $t = 0, 1, \ldots$, consider the situation where a sequence of $\mathbb{R}^q$-valued rvs $\{\Lambda^{(N)}(t),\ N = 1, 2, \ldots\}$ is weakly convergent as in (7.23) *together* with (4.30), *i.e.,* we have the *joint* convergence

$$\left(\sqrt{N}L_0^{(N)}(t), \Lambda^{(N)}(t)\right) \Longrightarrow_N (L_0(t), \Lambda(t)). \tag{7.27}$$

Then, it holds that

$$\left(\sqrt{N}Z_g^{(N)}(t), \Lambda^{(N)}(t)\right) \Longrightarrow_N \left(f'(q(t))R_g(t)L_0(t), \Lambda(t)\right). \tag{7.28}$$

7.5 A Proof of Proposition 3

Fix $t = 0, 1, \ldots$. Going back to the basic decomposition (7.18), we note that

$$L_g^{(N)}(t+1) = L_{\hat{g}_t}^{(N)}(t) + H_{\hat{g}}^{(N)}(t+1) + Z_{\hat{g}}^{(N)}(t) \tag{7.29}$$

for each $N = 1, 2, \ldots$, where $\hat{g}$ and $\hat{g}_t$ are the mappings $\mathbb{N} \to \mathbb{R}^p$ defined earlier at (7.15) and (7.17), respectively.

By Corollary 1, we already have

$$\sqrt{N}H_{\hat{g}}^{(N)}(t+1) \Longrightarrow_N H_{\hat{g}}(t+1) \tag{7.30}$$

where the rv $H_{\hat{g}}(t+1)$ is a zero-mean Gaussian rv with covariance matrix $\Sigma_{\hat{g}}(t+1)$. However, note that the rvs $L^{(N)}_{\hat{g}_t}(t)$ and $Z^{(N)}_{\hat{g}}(t)$ are both $\mathcal{F}_t$-measurable for each $N = 1, 2, \ldots$. Therefore, by the comments following Corollary 1, Proposition 3 will be established if we can show the *joint* convergence

$$\sqrt{N}(L^{(N)}_{\hat{g}_t}(t), Z^{(N)}_{\hat{g}}(t)) \Longrightarrow_N (L_{\hat{g}_t}(t), Z_{\hat{g}}(t)) \tag{7.31}$$

for some $\mathbb{R}^{2p}$-valued rv $(L_{\hat{g}_t}(t), Z_{\hat{g}}(t))$ with

$$Z_{\hat{g}}(t) = f'(q(t))R_g(t)L_0(t). \tag{7.32}$$

Indeed, as indicated at the end of Section 7.3, Equations (7.30) and (7.31) imply

$$\sqrt{N}(L^{(N)}_{\hat{g}_t}(t), Z^{(N)}_{\hat{g}}(t), H^{(N)}_{\hat{g}}(t+1)) \Longrightarrow_N (L_{\hat{g}_t}(t), Z_{\hat{g}}(t), H_{\hat{g}}(t+1)) \tag{7.33}$$

with the Gaussian rv $H_{\hat{g}}(t+1)$ independent of $(L_{\hat{g}_t}(t), Z_{\hat{g}}(t))$. Consequently, we have the convergence

$$\sqrt{N}L^{(N)}_g(t+1) \Longrightarrow_N L_{\hat{g}_t}(t) + Z_{\hat{g}}(t) + H_{\hat{g}}(t+1). \tag{7.34}$$

Combining (7.34) with (7.32) yields the desired result.

To establish the validity of (7.31) and (7.32), we proceed as follows: Under [**E:t**], it is already the case that

$$\sqrt{N}(L^{(N)}_0(t), L^{(N)}_{\hat{g}_t}(t)) \Longrightarrow_N (L_0(t), L_{\hat{g}_t}(t)) \tag{7.35}$$

while the strengthening (7.27)-(7.28) (with $\Lambda^{(N)}(t) = \sqrt{N}L^{(N)}_{\hat{g}_t}(t)$) of Proposition 5 leads to (7.28) in the form

$$\sqrt{N}\left(L^{(N)}_{\hat{g}_t}(t), Z^{(N)}_g(t)\right) \Longrightarrow_N (L_{\hat{g}_t}(t), f'(q(t))R_g(t)L_0(t))\,. \tag{7.36}$$

In other words, joint convergence (7.31) takes place with the identification (7.32) for the limiting rv.

7.6 A Proof of Proposition 1

Fix $t = 0, 1, \ldots$ and $N = 1, 2, \ldots$. We start with the observation that

$$\begin{aligned} & \sqrt{N}\left(f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right) \\ = \; & \sqrt{N}\int_{q(t)}^{\frac{Q^{(N)}(t)}{N}} (f'(x) - f'(q(t)))\, dx + \sqrt{N}\left(\frac{Q^{(N)}(t)}{N} - q(t)\right) f'(q(t)) \\ = \; & \sqrt{N}U^{(N)}(t) + \sqrt{N}L_0^{(N)}(t) f'(q(t)) \end{aligned} \tag{7.37}$$

where we have set

$$U^{(N)}(t) := \int_{q(t)}^{\frac{Q^{(N)}(t)}{N}} (f'(x) - f'(q(t)))\, dx.$$

The desired conclusion (4.31) will readily follow if we show the convergence

$$\sqrt{N}U^{(N)}(t) \stackrel{P}{\to}_N 0. \tag{7.38}$$

To that end, fix $\varepsilon > 0$ arbitrary. For any $\delta > 0$, we have

$$\begin{aligned} \mathbf{P}\left[\sqrt{N}\left|U^{(N)}(t)\right| > \varepsilon\right] \; \leq \; & \mathbf{P}\left[\sqrt{N}\left|U^{(N)}(t)\right| > \varepsilon,\ \left|\frac{Q^{(N)}(t)}{N} - q(t)\right| \leq \delta\right] \\ & + \mathbf{P}\left[\left|\frac{Q^{(N)}(t)}{N} - q(t)\right| > \delta\right]. \end{aligned} \tag{7.39}$$

By the continuity of f' at $x = q(t)$, we know that for each $\eta > 0$, there exists $\delta(\eta) > 0$ such that whenever $|x - q(t)| < \delta(\eta)$ in $\mathbb{R}_+$, we get

$$|f'(x) - f'(q(t))| \leq \eta.$$

Now fix $\eta > 0$ and pick $\delta > 0$ in the range $(0, \delta(\eta))$. Thus, on the event $[\left|\frac{Q^{(N)}(t)}{N} - q(t)\right| \leq \delta]$, we find that

$$\sqrt{N}\left|U^{(N)}(t)\right| \leq \sqrt{N}\eta\left|\frac{Q^{(N)}(t)}{N} - q(t)\right|.$$

Reporting this fact into the inequality (7.39) we obtain

$$\begin{aligned} & \mathbf{P}\left[\sqrt{N}\left|U^{(N)}(t)\right|>\varepsilon\right] \\ \leq\ & \mathbf{P}\left[\sqrt{N}\left|\frac{Q^{(N)}(t)}{N}-q(t)\right|>\frac{\varepsilon}{\eta}\right]+\mathbf{P}\left[\left|\frac{Q^{(N)}(t)}{N}-q(t)\right|>\delta\right] \end{aligned} \tag{7.40}$$

Letting N go to infinity in (7.40) and using the convergence (4.6) and (4.30), we get

$$\limsup_{N\to\infty}\mathbf{P}\left[\sqrt{N}\left|U^{(N)}(t)\right|>\varepsilon\right]\leq\mathbf{P}\left[L_0(t)>\frac{\varepsilon}{\eta}\right]. \tag{7.41}$$

The desired conclusion (7.38) is now immediate upon letting $\eta>0$ go to zero in this last inequality since its left-hand side is independent of η.

7.7 Strengthening Claim (ii) of Theorem 2

Before turning to the proof of Proposition 4 in the next section, we pause to present a result that builds on Claim (ii) of Theorem 2. This result will prove useful in establishing Proposition 4 later on.

Proposition 6. *Assume (AW1)-(AW2) to hold. Then, for each $t=0,1,\ldots$, and any function $g:\mathbb{N}\to\mathbb{R}$, it holds that*

$$\frac{1}{N}\sum_{i=1}^{N}g(W_i^{(N)}(t))Z_i^{(N)}(t)^{\ell}\overset{P}{\to}_N\mathbf{E}\left[g(W(t))Z(t)^{\ell}\right] \tag{7.42}$$

for each integer $\ell=1,2,\ldots$.

Proof. Fix $t=0,1,\ldots$ and $\ell=1,2,\ldots$. Also fix $N=1,2,\ldots$ and $i=1,\ldots,N$. We have

$$\begin{aligned} & g(W_i^{(N)}(t))Z_i^{(N)}(t)^{\ell} \\ =\ & g(W_i^{(N)}(t))\left(\left(1-f\left(\frac{Q^{(N)}(t)}{N}\right)\right)^{\ell W_i^{(N)}(t)}-\left(1-f(q(t))\right)^{\ell W_i^{(N)}(t)}\right) \\ & +g_{t,\ell}(W_i^{(N)}(t)) \end{aligned} \tag{7.43}$$

with mapping $g_{t,\ell} : \mathbb{R} \to \mathbb{R}$ given by

$$g_{t,\ell}(w) := g(w) \left(1 - f(q(t))\right)^{\ell w}, \quad w \in \mathbb{N}.$$

On the other hand, for any pair a, b in $[0, 1]$, we have

$$|a^p - b^p| = p| \int_a^b t^{p-1} dt| \leq p|b - a| \tag{7.44}$$

for each $p = 1, 2, \ldots$, so that

$$\begin{aligned} & \left| \left(1 - f\left(\frac{Q^{(N)}(t)}{N}\right)\right)^{\ell W_i^{(N)}(t)} - \left(1 - f(q(t))\right)^{\ell W_i^{(N)}(t)} \right| \\ \leq \quad & \ell W_i^{(N)}(t) \left| f\left(\frac{Q^{(N)}(t)}{N}\right) - f(q(t)) \right|. \end{aligned} \tag{7.45}$$

Therefore,

$$\begin{aligned} & |\frac{1}{N} \sum_{i=1}^N g(W_i^{(N)}(t)) \left(Z_i^{(N)}(t)^\ell - \left(1 - f(q(t))\right)^{\ell W_i^{(N)}(t)} \right)| \\ \leq \quad & \ell \left| f\left(\frac{Q^{(N)}(t)}{N}\right) - f(q(t)) \right| \left(\frac{1}{N} \sum_{i=1}^N W_i^{(N)}(t) |g(W_i^{(N)}(t))| \right) \end{aligned} \tag{7.46}$$

and using the convergence (7.3), we get

$$\frac{1}{N} \sum_{i=1}^N g(W_i^{(N)}(t)) \left(Z_i^{(N)}(t)^\ell - \left(1 - f(q(t))\right)^{\ell W_i^{(N)}(t)} \right) \xrightarrow{P}_N 0 \tag{7.47}$$

since

$$\frac{1}{N} \sum_{i=1}^N W_i^{(N)}(t) |g(W_i^{(N)}(t))| \xrightarrow{P}_N \mathbf{E}\left[W(t)|g(W(t))|\right] \tag{7.48}$$

by Claim (ii) of Theorem 2.

The conclusion is now immediate from the decomposition (7.43), the convergence (7.48) and the convergence

$$\frac{1}{N} \sum_{i=1}^N g_{t,\ell}(W_i^{(N)}(t)) \xrightarrow{P}_N \mathbf{E}\left[g_{t,\ell}(W(t))\right] \tag{7.49}$$

obtained from Claim (ii) of Theorem 2. □

7.8 A Proof of Proposition 4

The proof of Proposition 4 relies on the following two technical lemmas; their proofs are omitted in the interest of brevity.

Lemma 6. *For any x in $\mathbb{R}$, the Taylor series expansion*

$$e^{jx} = 1 + jx - \frac{x^2}{2} + R(x) \tag{7.50}$$

holds, with complex-valued remainder term $R(x)$ satisfying

$$|R(x)| \leq \frac{|x|^3}{6}. \tag{7.51}$$

Lemma 7. *Consider the array of complex-valued rvs $\{C_{N,i},\ i = 1, \ldots, N;\ N = 1, 2, \ldots\}$ with $|C_{N,i}| < 1$ for $i = 1, \ldots, N$. If $\max_{i=1,\ldots,N} |C_{N,i}| \to_N 0$ a.s. and $\sum_{i=1}^N C_{N,i} \stackrel{P}{\to}_N \lambda$, then*

$$\prod_{i=1}^N (1 - C_{N,i}) \stackrel{P}{\to}_N e^{-\lambda}. \tag{7.52}$$

The proof of Proposition 4 can now proceed: Fix $N = 1, 2, \ldots$ and θ arbitrary in $\mathbb{R}$. By conditional independence, we find that

$$\begin{aligned} & \mathbf{E}\left[\exp\left(j\theta\sqrt{N}H_g^{(N)}(t+1)\right)|\mathcal{F}_t\right] \\ = & \prod_{i=1}^N \mathbf{E}\left[\exp\left(j\frac{\theta}{\sqrt{N}}\left(g(W_i^{(N)}(t))\left(M_i^{(N)}(t+1) - Z_i^{(N)}(t)\right)\right)\right)|\mathcal{F}_t\right] \\ = & \prod_{i=1}^N \left(1 - C_i^{(N)}(t)\right) \end{aligned} \tag{7.53}$$

with

$$\begin{aligned} C_i^{(N)}(t) = & Z_i^{(N)}(t)\left[1 - \exp\left(j\frac{\theta}{\sqrt{N}}g(W_i^{(N)}(t))(1 - Z_i^{(N)}(t))\right)\right] \\ & + (1 - Z_i^{(N)}(t))\left[1 - \exp\left(-j\frac{\theta}{\sqrt{N}}g(W_i^{(N)}(t))Z_i^{(N)}(t)\right)\right] \end{aligned} \tag{7.54}$$

for each $i = 1, \ldots, N$.

In view of these remarks, the desired result (7.19) is now a simple consequence of Lemma 7 provided the required conditions can be shown hold, namely

$$\lim_N \max_{i=1,\ldots,N} |C_i^{(N)}(t)| = 0 \quad a.s. \tag{7.55}$$

and

$$\sum_{i=1}^{N} C_i^{(N)}(t) \xrightarrow{P}_N \sigma_g(t+1) \tag{7.56}$$

with $\sigma_g(t+1)$ given by (7.20).

Condition (7.55) trivially holds. To establish (7.56) we invoke Lemma 6 to write

$$C_i^{(N)}(t) = Z_i^{(N)}(t) B_{i,1}^{(N)}(t) + (1 - Z_i^{(N)}(t)) B_{i,2}^{(N)}(t) \tag{7.57}$$

with

$$\begin{aligned} B_{i,1}^{(N)}(t) &= -j\frac{\theta}{\sqrt{N}} g(W_i^{(N)}(t))(1 - Z_i^{(N)}(t)) \\ &\quad + \frac{\theta^2}{2N} g(W_i^{(N)}(t))^2 (1 - Z_i^{(N)}(t))^2 + \beta_{i,1}^{(N)}(t;\theta) \end{aligned} \tag{7.58}$$

and

$$\begin{aligned} B_{i,2}^{(N)}(t) &= j\frac{\theta}{\sqrt{N}} g(W_i^{(N)}(t)) Z_i^{(N)}(t) \\ &\quad + \frac{\theta^2}{2N} g(W_i^{(N)}(t))^2 Z_i^{(N)}(t)^2 + \beta_{i,2}^{(N)}(t;\theta) \end{aligned} \tag{7.59}$$

where remainder terms $\beta_{i,1}^{(N)}(t;\theta)$ and $\beta_{i,2}^{(N)}(t;\theta)$ in these expressions satisfy

$$|\beta_{i,1}^{(N)}(t;\theta)|, |\beta_{i,2}^{(N)}(t;\theta)| \leq C\frac{|\theta|^3}{6\sqrt{N^3}}$$

for some positive constant C, say $C := \max\{|g(w)|^3, w = 1, \ldots, W_{\max}\}$.

Reporting (7.58) and (7.59) into (7.57) and simplifying the resulting expression, we find

$$C_i^{(N)}(t) = \frac{\theta^2}{2N} g(W_i^{(N)}(t))^2 Z_i^{(N)}(t)(1 - Z_i^{(N)}(t)) + \gamma_i^{(N)}(t;\theta) \tag{7.60}$$

with remainder term

$$\gamma_i^{(N)}(t;\theta) := Z_i^{(N)}(t)\beta_{i,1}^{(N)}(t;\theta) + (1 - Z_i^{(N)}(t))\beta_{i,2}^{(N)}(t;\theta).$$

Obviously,

$$|\gamma_i^{(N)}(t;\theta)| \leq C\frac{|\theta|^3}{6\sqrt{N^3}}. \tag{7.61}$$

It is now plain that

$$\sum_{i=1}^{N} C_i^{(N)}(t) = \frac{\theta^2}{2}\left(\frac{1}{N}\sum_{i=1}^{N} g(W_i^{(N)}(t))^2 Z_i^{(N)}(t)(1 - Z_i^{(N)}(t))\right) + \Gamma^{(N)}(t) \tag{7.62}$$

where we have set

$$\Gamma^{(N)}(t) = \frac{1}{N}\sum_{i=1}^{N}\gamma_i^{(N)}(t;\theta).$$

By Proposition 6 (with $\ell = 1$ and $\ell = 2$) we get

$$\frac{1}{N}\sum_{i=1}^{N} g(W_i^{(N)}(t))^2 Z_i^{(N)}(t)(1 - Z_i^{(N)}(t)) \xrightarrow{P}_N \sigma_g(t+1) \tag{7.63}$$

with $\sigma_g(t+1)$ given by (7.20), while $\lim_{N\to\infty}\Gamma^{(N)}(t) = 0$ a.s. by virtue of (7.61). The desired conclusion (7.56) is obtained and the proof of Proposition 4 is complete.

7.9 A Proof of Proposition 5

Fix $N = 1, 2, \ldots$ and $i = 1, \ldots, N$. Observe that

$$\begin{aligned}
& \gamma^{(N)}(t)^{W_i^{(N)}(t)} - \gamma(t)^{W_i^{(N)}(t)} \\
= \;& W_i^{(N)}(t)\int_{f(q(t))}^{f(\frac{Q^{(N)}(t)}{N})} (1-y)^{W_i^{(N)}(t)-1}\; dy \\
= \;& W_i^{(N)}(t)\Delta_i^{(N)}(t) \\
& + W_i^{(N)}(t)\left(f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right)(1 - f(q(t)))^{W_i^{(N)}(t)-1}
\end{aligned} \tag{7.64}$$

where we have set

$$\Delta_i^{(N)}(t) := \int_{f(q(t))}^{f(\frac{Q^{(N)}(t)}{N})} \left[(1-y)^{W_i^{(N)}(t)-1} - (1-f(q(t)))^{W_i^{(N)}(t)-1}\right]\, dy.$$

Consequently,

$$\begin{aligned}\sqrt{N}Z_g^{(N)}(t) &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} g(W_i^{(N)}(t))W_i^{(N)}(t)\Delta_i^{(N)}(t)\right) \\ &\quad + \left(\frac{1}{N}\sum_{i=1}^{N} g_t^\star(W_i^{(N)}(t))\right)\cdot\sqrt{N}\left(f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right)\end{aligned} \tag{7.65}$$

where the mapping $g_t^\star : \mathbb{N} \to \mathbb{R}^p$ is defined by

$$g_t^\star(w) := wg(w)\left(1-f(q(t))\right)^{w-1}, \quad w \in \mathbb{N}.$$

Using the inequality (7.44), we find that

$$\begin{aligned}|\Delta_i^{(N)}(t)| &\leq \left(W_i^{(N)}(t)-1\right)\left|\int_{f(q(t))}^{f(\frac{Q^{(N)}(t)}{N})} |y-f(q(t))|\, dy\right| \\ &\leq W_{\max}\left|\int_{f(q(t))}^{f(\frac{Q^{(N)}(t)}{N})} |y-f(q(t))|\, dy\right| \\ &= \frac{1}{2}W_{\max}\left|f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right|^2\end{aligned} \tag{7.66}$$

Thus,

$$\begin{aligned}&\sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^{N} g(W_i^{(N)}(t))W_i^{(N)}(t)\Delta_i^{(N)}(t)\right\| \\ \leq\ &\frac{1}{2}W_{\max}\sqrt{N}\left|f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right|^2\left(\frac{1}{N}\sum_{i=1}^{N}\|g(W_i^{(N)}(t))\|W_i^{(N)}(t)\right).\end{aligned}$$

Let N go to infinity. Coupling the convergence (7.3) with Proposition 1 we find that

$$\sqrt{N}\left|f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right|^2 \xrightarrow[N]{P} 0.$$

It now follows by Claim (ii) of Theorem 2 (applied to the mapping $w \to w|g(w)|$) that

$$\sqrt{N}\left\|\frac{1}{N}\sum_{i=1}^{N} g(W_i^{(N)}(t))W_i^{(N)}(t)\Delta_i^{(N)}(t)\right\| \stackrel{P}{\to}_N 0. \tag{7.67}$$

Thus, in view of the decomposition (7.64) (with (7.65)) and of the convergence (7.67), it is now plain that the convergence (7.25) will hold provided we can show that

$$\sqrt{N}\left(f(\frac{Q^{(N)}(t)}{N}) - f(q(t))\right)\left(\frac{1}{N}\sum_{i=1}^{N} g_t^\star(W_i^{(N)}(t))\right) \Longrightarrow_N f'(q(t))R_g(t)L_0(t). \tag{7.68}$$

However,

$$\frac{1}{N}\sum_{i=1}^{N} g_t^\star(W_i^{(N)}(t)) \stackrel{P}{\to}_N R_g(t)$$

and (7.68) follows by applying Proposition 1.

Chapter 8

Conclusions and Future Directions

8.1 Conclusions

The main focus of this dissertation is on characterizing the dynamics of the TCP congestion-control mechanism and AQM (in particular RED) in the regime where the number of competing flows is large. The systems we considered can be thought of as stochastic feedback systems. The first system we considered is the rate-based model in Chapter 3 and the Weak Law of Large Numbers (Theorem 1) is established for the model. It reveals the natural simplification of the system dynamics. In fact, Theorem 1 indicates that the asymptotic queue follows a simple deterministic recursion which involves only the expectation of the limiting number of packets injected into the network in a timeslot. The recursion for the limiting number of packets injected into the network is also closely related to the recursion of a single traffic flow, *i.e.,* at any time they both have the same conditional expectation given the same state in the previous timeslot and marking/dropping probability.

While the model is subsequently refined to be more realistic in Chapter 4 – 5, similar results can also be established. These Weak Laws of Large Numbers

therefore provide justification for the analysis of the TCP/AQM interaction through a deterministic feedback system when there exists a large number of flows. More specifically, the recursion of the asymptotic (or average) queue, *i.e.,* $q(t)$, depends only on the capacity of the queue and the *expected* amount of traffic injected into the network in each timeslot, *i.e.,* $\mathbf{E}[A(t)]$. This is similar to many of the existing models which analyze the dynamics of TCP and AQM mechanisms as deterministic feedback-delay systems.

Following the aforementioned observation, this research complements the control-theoretic studies of TCP/AQM dynamics. Much of the research effort along this direction has been focused on either the locally or globally stability of the controlled queue (as surveyed in Section 2.2). These work reveal sufficient conditions for the queue to either be local or global asymptotically stable, *i.e.,* the queue eventually settles to an equilibrium value. Under these sufficient conditions, Assumption (AW4), *i.e.,* $(q(t), \mathbf{Y}(t)) \Longrightarrow_N (q^\star, \mathbf{Y}^\star)$, is reasonable. In other words, the average behavior of such complex stochastic feedback system at the equilibrium becomes easy to specify through steady-state analysis given that there is a large number of flows and the system parameters are set appropriately using the sufficient conditions from the control-theoretic analysis.

The system in the large number of flows asymptotic regime at the steady-state is simple to analyze because of the behavior of any single flow no longer influences the network behavior. Also, the finding that the flows become asymptotically independent also supports the belief that RED breaks global synchronization with a large number of flows.

As expected, the natural simplification of the models and these asymptotic properties can be found in the regime where the number of interacting flows is

large without having to rely on ad-hoc assumptions during the analysis. Furthermore, the limiting models are also compatible with other previously proposed models in their respective regime, thereby revealing the versatility of the limiting models.

While the Weak Laws of Large Numbers capture the average behavior of the systems, the Central Limit analysis capture the errors/uncertainties due to the randomness in the system. These uncertainties appear as fluctuations in the queue and are not represented in the deterministic models.

It is noteworthy from the Central Limit analysis that the random marking mechanism in AQM always introduces random fluctuations in the queue. This is an intrinsic behavior for the random marking mechanism due to the limited granularity in the feedback information. As noted in Chapter 5, such a fluctuation can be reduced by improving the quality of the feedback information either through an increase in the number of ECN bits or through an in-band signaling mechanism as suggested in [56]. There are, however, other means to avoid such fluctuations. For example, TCP Vegas [9] and its variants such as TCP FAST [27] use delay information instead of marks to adjust their congestion windows. Given an accurate timestamp in the packets, each flow can accurately calculate the appropriate adjustment to its congestion window size, thus greatly reducing uncertainty from the randomness in the marks.

Although the fluctuations due to randomness can be well-approximated through the CLT, the feedback system in the deterministic (or limiting) models also introduces fluctuations (or oscillations) as classical control-theoretic analysis indicates. The random fluctuations derived in the Central Limit analysis complement these oscillations. Moreover, some parameters in the system can also

induce fluctuations in both components, *e.g.,* the steep slope of the marking probability function not only causes fluctuation in the random components as suggested by the Delta Method but also induces oscillations as the feedback gain being large in the deterministic model.

Overall, this Ph.D. dissertation provides a novel framework for characterizing the asymptotic dynamics of the network containing many flows where flows interacts with each other through the random feedback signaling from the network. The results obtained are both compatible with and complementary to the existing literature. The approach used here can also be generalized to other weakly interacting stochastic feedback systems.

8.2 Future Directions

While this thesis studies in depth the problem of characterizing the dynamics of the TCP congestion-control mechanism and AQM schemes, there remain several open problems for research as outlined below.

Uncontrolled cross traffic

First, consider the situation where the bottleneck gateway has uncontrolled cross traffic, *i.e.,* the amount of traffic injected into the network is unresponsive to the network congestion, then the overall number of packets arriving at the bottleneck gateway will be a combination from both TCP flows and uncontrolled flows. If the queue is small and the amount of uncontrolled cross traffic is small comparing to the capacity (as it should in any stable network), then a reasonable approximation is to reduce the capacity of the bottleneck gateway equal to the

number of packets from the uncontrolled flows in each timeslot in the model. The only difference in the modified model from the models considered in this thesis is that the available capacity in the bottleneck gateway varies as a function of time. In [52], a discrete-time traffic model such as $M/G/\infty$ traffic arrival process has been found suitable for modeling the uncontrolled cross traffic comprising of many uncontrolled flows.

Multiple bottleneck

While recent work has shown that a single bottleneck model is a reasonable approximation for an analysis of open-loop or uncontrolled flows in a multi-stage network of queues [12], it is not clear that this would be the case for feedback-based flows such as TCP. Studies have shown that a large number of TCP flows over multiple bottleneck queue utilizing Tail Drop gateway exhibits very complex behavior [3]. It remains an open problem on how a collection of TCP flows will behave in a general network setting. Deterministic models of congestion-control utilizing an optimization framework are proposed in [30, 37]. It merits an investigation as to whether these deterministic models can be justified under a large number of flows asymptotic similar to the single bottleneck case.

Wireless physical layer

In all of the models considered in this thesis, the only point of interaction between flows are at the bottleneck gateway. However, if flows also share a common wireless physical layer (such as the third generation wireless data networks or ad-hoc networks), then flows can also interact in the physical/MAC layer as well. For example, the flows utilizing the downlink of the cdma2000

1xEV-DO (IS-856) wireless cellular network will interact through the *proportional fair* scheduler and their uplink packets (*e.g.,* ACK packets) will interact through interference in the physical layer and its MAC layer mechanism [62]. The effect of this added layer of interaction to the system is difficult to analyze and quantify using traditional modeling techniques. Nevertheless, we can expect that model simplification will occur in the large number of flows asymptotic regime. This might shed some insights on the complex behavior of the system.

Selection of the feedback function

As previously mentioned in the discussion on the CLT in Chapter 5, the selection of the appropriate feedback mechanism can now be evaluated systematically through the limiting models.

Extension to other congestion controllers and AQM mechanisms

Although the modeling efforts in this thesis are concentrated on TCP/RED which are the dominant combination of the existing congestion-control and AQM mechanism, it is easy to generalize the approach to other congestion controllers and AQM mechanisms (see [58] for an example). We will briefly describe such generalizations in this section.

AQM gateways control their level of congestion by randomly marking incoming packets to signal the traffic sources of the congestion level. In order to do so, each AQM mechanism calculates a quantity which can be referred to as a *congestion measure*. The congestion measure in a timeslot can be a function of the queue sizes (current size and past values), of the volume of the incoming traffic, and of the previous values of the congestion measure. We can then use a

continuous mapping to map this congestion measure to the marking probability in the AQM mechanism. Under appropriate scaling rules, this marking probability will converge under a large number of flows.

The congestion controllers can be generalized through the deterministic mappings which map the current and past history of the state variables (*e.g.,* congestion window) and marks from the AQM mechanisms to the state variables in the next timeslot. Under appropriate assumptions, the limiting model of the generalized system can be derived along the lines of the analysis given in this dissertation.

Bibliography

[1] Cdric Adjih, Philippe Jacquet, and Nikita Vvedenskaya. Performance evaluation of a single queue under multi-user TCP/IP connections. Technical Report RR-4141, INRIA, March 2001.

[2] E. Altman, K. Avrachenkov, and C. Barakat. TCP in presence of bursty losses. In *Proceedings of SIGMETRICS*, 2000.

[3] Francois Baccelli and Dohy Hong. Interaction of TCP flows as billiards. In *Proceedings of IEEE INFOCOM*, San Francisco, CA, 2003.

[4] Francois Baccelli, Dohy Hong, and Zhen Liu. Fixed point methods for the simulation of the sharing of a local loop by a large number of interacting TCP connections. Technical Report RR-4154, INRIA, April 2001.

[5] Francois Baccelli, David R. McDonald, and Julien Reynier. A mean-field model for multiple TCP connections through a buffer implementing RED. Technical report, INRIA, April 2002.

[6] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York (NY), 1968.

[7] Thomas Bonald. Comparison of TCP Reno and TCP Vegas via fluid approximation. Technical Report 3563, INRIA, November 1998.

[8] Thomas Bonald, M. May, and J. Bolot. Analytic evaluation of RED performance. In *Proceedings of IEEE INFOCOM*, 2000.

[9] Lawrence S. Brakmo and Larry L. Peterson. TCP Vegas: end to end congestion avoidance on a global internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–1480, October 1995.

[10] Arjan Durresi, Mukundan Sridharan, Chunlei Liu, Mukul Goyal, and Raj Jain. Multilevel early congestion notification. In *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics*, pages 12–17, Orlando, FL, July 2001.

[11] A.K. Erlang. The theory of probabilities and telephone conversation. *Nyt Tidsskrift Mat.*, 1909.

[12] D. Eun and Ness B. Shroff. Simplification of network analysis in large-bandwidth systems. In *Proceedings of IEEE INFOCOM*, San Francisco, CA, 2003.

[13] W. Feng, D. Kandlur, D. Saha, and K. Shin. The blue queue management algorithms. *IEEE/ACM Transactions on Networking*, 10(4), August 2002.

[14] Victor Firoiu and Marty Borden. A study of active queue management for congestion control. In *Proceedings of IEEE INFOCOM*, 2000.

[15] Sally Floyd. TCP and explicit congestion notification. *Computer Communication Review*, 24(5):10–23, October 1994.

[16] Sally Floyd and Kevin Fall. Router mechanisms to support end-to-end congestion control. Technical report, Lawrence Berkeley National Lab, 1997.

[17] Sally Floyd, R. Gummadi, and Scott Shenker. Adaptive RED: an algorithm for increasing the robustness of RED Active Queue Management. Submitted for publication, August 2001.

[18] Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1995.

[19] Michele Garetto, Renato Lo Cigno, Michela Meo, and Marco Ajmone Marsan. A detailed and accurate closed queueing network model of many interacting TCP flows. In *Proceedings of IEEE INFOCOM*, pages 1706–1715, 2001.

[20] Fred Halsall. *Data Communications, Computer Networks and Open Systems*. Addison-Wesley, 1997.

[21] G. Hasegawa and M. Murata. Analysis of dynamic behaviors of many TCP connections sharing Tail–Drop/RED routers. In *Proceedings of IEEE GLOBECOM*, November 2001.

[22] C. Hollot, V. Misra, D. Towsley, and W. Gong. A control theoretic analysis of RED. In *Proceedings of IEEE INFOCOM*, 2001.

[23] C. V. Hollot, Yong Liu, Vishal Misra, and Don Towsley. Unresponsive flows and AQM performance. In *Proceedings of IEEE INFOCOM*, April 2003.

[24] C. V. Hollot, Vishal Misra, Donald F. Towsley, and Weibo Gong. On designing improved controllers for AQM routers supporting TCP flows. In *Proceedings of IEEE INFOCOM*, pages 1726–1734, 2001.

[25] Dohy Hong and Dmitri Lebedev. Many TCP user asymptotic analysis of the AIMD model. Technical Report RR-4229, INRIA, July 2001.

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

[27] Cheng Jin, David X. Wei, and Steven H. Low. FAST TCP: motivation, architecture, algorithms, performance. Submitted for publication.

[28] Ramesh Johari and David Kim Hong Tan. End-to-end congestion control for the Internet: Delays and stability. *IEEE/ACM Transactions on Networking*, 9(6):818–832, December 2001.

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

[30] Frank Kelly, A.K. Maulloo, and D.K.H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[31] Frank P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.

[32] Arzad Kherani and Anurag Kumar. Stochastic models for throughput analysis of randomly arriving elastic flows in the Internet. In *Proceedings of IEEE INFOCOM*, New York City, NY, July 2002.

[33] A.Y. Khintchine. Mathematisches uber die erwortung vor einemoffenthchen schalter. *Mat. Sb.*, 1932.

[34] S. Kunniyur and R. Srikant. Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management. In *Proceedings of ACM SIGCOMM*, San Diego, August 2001.

[35] Richard J. La, Priya Ranjan, and Eyad H. Abed. Analysis of ARED. In *Proceedings of the 18th International Teletraffic Congress*, Berlin, Germany, September 2003.

[36] Steven H. Low. A duality model of TCP and queue management algorithms. In *Proceedings of ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management*, September 2000.

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

[38] 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*, New York City, NY, July 2002.

[39] D.-J. Ma, Armand M. Makowski, and A. Shwartz. Stochastic approximations for finite–state Markov chains. *Stochastic Processes and Their Applications*, 35:27–45, 1990.

[40] M. Mathis, J. Semske, J. Mahdavi, and T. Ott. The macroscopic behavior of TCP congestion avoidance algorithm. *Computer Communication Review*, 27(3), July 1997.

[41] M. May, J. Bolot, C. Diot, and B. Lyles. Reasons not to deploy RED. In *Proceedings of 7th. International Workshop on Quality of Service (IWQoS'99)*, June 1999.

[42] M. Mellia, I. Stoica, and H. Zhang. TCP model for short lived flows. *IEEE Communications Letters*, 6(2):85–87, 2002.

[43] V. Misra, W. Gong, and D. Towsley. Stochastic differential equation modeling and analysis of TCP windowsize behavior. In *Proceedings of Performance*, 1999.

[44] Jeonghoon Mo, Richard J. La, Venkat Anantharam, and Jean Walrand. Analysis and comparison of TCP Reno and Vegas. In *Proceedings of IEEE INFOCOM*, 1999.

[45] UCB/VINT/LBNL Network Simulator (ns) version 2.0.

[46] Teunis J. Ott, T.V. Lakshman, and Larry Wong. SRED: Stabilized RED. In *Proceedings of IEEE INFOCOM*, 1999.

[47] J. Padhye, V. Firoiu, and D. Towsley. A stochastic model of TCP Reno congestion avoidance and control. Technical Report 99-02, Department of Computer Science, University of Massachusetts, Amherst, 1999.

[48] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP Reno performance: A simple model and its empirical validation. *IEEE/ACM Transactions on Networking*, April 2000.

[49] Fernando Paganini. A global stability result in network flow control. *Systems and Control Letters*, 46(6):153–163, 2002.

[50] C. Palm. Analysis of the Erlang traffic formulae for busy-signal arrangements. *Ericsson Tech.*, 1938.

[51] Rong Pan, Balaji Prabhakar, and Konstantinos Psounis. CHOKe: a stateless active queue management scheme for approximating fair bandwidth allocation. In *Proceedings of IEEE INFOCOM*, volume 2, pages 942–951, Tel Aviv, Israel, March 2000.

[52] Minothi Parulekar and Armand M. Makowski. M/G/infinity input processes: A versatile class of models for network traffic. In *Proceedings of IEEE INFOCOM*, pages 419–426, 1997.

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

[54] Priya Ranjan, Eyad H. Abed, and Richard J. La. Communication delay and instability in rate-controlled networks. In *Proceedings of IEEE Conference on Decision and Control*, 2003.

[55] Sanjay Shakkottai and R. Srikant. How good are deterministic fluid models of Internet congestion control? In *Proceedings of IEEE INFOCOM*, New York, NY, June 2002.

[56] Mayank Sharma, Dina Katabi, Balaji Prabhakar, and Rong Pan. A general multiplexed ECN channel and its use for wireless loss notification. Submitted for Publication, February 2003.

[57] Peerapol Tinnakornsrisuphap and Richard J. La. Asymptotic behavior of heterogeneous TCP flows and RED gateways. Technical report, Institute for Systems Research, University of Maryland, 2003.

[58] Peerapol Tinnakornsrisuphap and Richard J. La. Characterization of queue fluctuations in probabilistic AQM mechanisms. In *Proceedings of ACM SIGMETRICS*, New York City, New York, June 2004.

[59] Peerapol Tinnakornsrisuphap and Armand Makowski. Queue dynamics of RED gateways under large number of TCP flows. In *Proceedings of IEEE GLOBECOM*, 2001.

[60] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[61] Andras Veres and Miklos Boda. The chaotic nature of TCP congestion control. In *Proceedings of INFOCOM'2000*, Tel-aviv, Israel, 2000.

[62] Qiang Wu and Eduardo Esteves. The cdma2000 high rate packet data system. In Jiangzhou Wang and Tung-Sang Ng, editors, *Advances in 3G Enhanced Technologies for Wireless Communications*. Artech House, 2002.

[63] L. Zhang and D. Clark. Oscillating behavior of network traffic: A case study simulation. *Internetworking: Research and Experience*, 1(2):101–112, 1990.

[64] 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 133–145, September 1991.