**UNIVERSITY OF SOUTHAMPTON**

Faculty of Engineering, Science and Mathematics

School of Electronics and Computer Science

A project report submitted for the award of

Master of Science in Radio Frequency Communication Systems

Supervisors: Professor Eric Rogers,

Dr. Stephan Weiss

# Congestion Control of Ad Hoc Wireless LANs: A Control-theoretic paradigm to digital filter based solution

by Rishad Ahmed Shafik

September 29, 2005

**Congestion Control of Ad Hoc Wireless LANs: A Control-theoretic paradigm to digital filter based solution**

## ABSTRACT

An ad hoc wireless LAN is a collection of wireless mobile nodes dynamically forming a temporary network without the use of any pre-existing network infrastructure or centralized administration. Due to its distributed nature, flexibility, robustness and ease of installation, ad hoc wireless LAN has greatly increased the scope for research in wireless communications. Since there is no defined structure, congestion control for systems where each ad hoc node can request certain bandwidth can pose the challenge of uncertain delay and instability and thus remains as a challenge in research. An ideal congestion control scheme for multi-hop ad hoc network would have to ensure that the bandwidth requests and input and output rates are regulated from chosen bridges and also from source and destination controllers. In this thesis, a novel congestion control scheme for multihop wireless LAN based on time-delay model is developed. The design of the proposed control model is derived from internal model control principles, with the control being done by the model reference controller and the error controller. Based on the congestion scenarios, the reference controller sets up a feasible reference value for the queue length, while the error controller feeds back rate-based compensation for the error between the reference and instantaneous queue lengths to combat against congestive disturbances. The proposed scheme makes use of Smith Predictor in the error controller to compensate for backward delay time, which is often referred to as "dead time" in control-engineering terms, to mitigate the stability problems that may occur. Underpinning the continuous-time model, a discretized and simplified digital-filter based solution is devised to make use of fast digital-filters available to date, without causing problem to scalability of the rate-based scheme and to propose a hardware based solution. The control objectives will be set with an aim to ensure full-link utilization and to achieve maximum rate recovery as soon as the congestion has been cleared under system stability. Simulations are performed to illustrate the performance of the controller under different congestion scenarios.

# Contents

# List of Figures

# List of Tables

# List of Symbols

| | |
|---|---|
| $r_{\mathrm{MAX}}(t)$ | Maximum throughput rate in bps |
| $r_{\mathrm{IN}}(t)$ | Input rate in bps |
| $r_{\mathrm{OUT}}(t)$ | Output rate in bps |
| $r_{\mathrm{DES}}(t)$ | Desired input rate in bps |
| $r_{\mathrm{REQ}}(t)$ | Required output rate in bps |
| $r_{\mathrm{DES_{REF}}}(t)$ | Desired input reference rate in bps |
| $r_{\mathrm{REQ_{REF}}}(t)$ | Required output reference rate in bps |
| $r_{\mathrm{DES_{ERR}}}(t)$ | Desired input rate error in bps |
| $r_{\mathrm{REQ_{ERR}}}(t)$ | Required input rate error in bps |
| $q(t)$ | Instantaneous queue length in bits |
| $q_{\mathrm{REF}}(t)$ | Reference queue length in bits |
| $d_{\mathrm{BW}}(t)$ | Backward disturbance rate in bps |
| $d_{\mathrm{FW}}(t)$ | Forward disturbance rate in bps |
| $T_{\mathrm{BW}}$ | Backward loop delay in seconds |
| $T_{\mathrm{FW}}$ | Forward loop delay in seconds |
| $T_{\mathrm{K}}$ | Buffer filling allowance time in seconds |

# Acknowledgements

This thesis is a result of research of four months and this is by far the most significant scientific accomplishment in my life. It would be impossible without support and appreciation of those who mattered most.

First of all, I would like to extend my thankful gratitude to my supervisor, Professor Eric Rogers. He has not only given me the impetus to work with this particular topic but also has helped me with his valuable advices and suggestions throughout the work. He envisioned that this could be one of the promising fields where classical control system can really come into play and believed that it could further open the door to ground-breaking research in congestion control of ad hoc wireless networks. His continued encouragement and excitement inpired me in the most important moments of making right decisions and I am glad to have worked with him. Without his guidance and advices, it would have been difficult to finish the work within this time.

I also sincerely thank my co-supervisor Dr. Stephan Weiss for giving me more than assistance throughout this work. His comments, corrections and suggestions have paved the way to a systematic approach to writing and critical thinking. I am particularly thankful to him for his generous demonstrations of LaTeX. Finally, I would like to thank him for supporting me during the 2nd IEE-EURASIP conference DSPenabledRadio, of which he was the general chair-person.

Last but not the least, there are those, whose support has been the key to continued spirit of work hard for this work. I sincerely thank my wife, my parents, my friends and my fellow students for their kind appreciation.

*To my daughter "Raya-moni",*
*my new inspiration of life ...*

# Chapter 1

# Introduction

In this chapter, brief definitions are presented for the terms used extensively in the forthcoming chapters and sections. Based on these definitions, the rest of the chapter will focus on the problem. An overview of this thesis is presented at the end of the chapter, together with brief summary of the rest of the chapters.

## 1.1 Wireless LANs

A network is a group of devices/nodes (*viz.* computers, mobile stations etc.) connected by a communication channel, capable of sharing information and other resources among themselves. A network can range from a peer-to-peer network connecting a small number of users in an office or department, to a local area network (LAN) connecting many users over permanently installed cables and dial-up lines, to a municipal area network or wide area network connecting users on several networks spread over a wide geographic area [10]. Networks can either be established over a wireless or wire-line channel. In wireless networks, a group of nodes are connected among themselves using technology other than conventional cables. These technologies include infrared line-of-sight high frequency light-wave signals for medium distance communication, high-frequency radio wave signals for short to long distance communication and spread spectrum signals for long distance communication. Since wireless local area network (WLAN) can provide mobility for its nodes, it is often chosen for personal communication devices and other portable communication devices. Depending on the distance of the wireless node from the network access-point, the communication speed can vary from $1 Mbps$ to several decades of $Mbps$ [16]. Wireless LANs are not always completely wireless and may be used to replace the cabling on certain network segments or to connect groups of networks that use conventional cabling. Similar to wire line networks, the nodes in the wireless LANs also be distinct depending on their role in the network. Some nodes act as client nodes, some as server or master nodes, while some nodes act as bridges, switches and hubs etc. In this thesis from this point forward, stations, terminals will only be termed as nodes to refer to smallest communication unit in the network. In the following sub-sections, a brief definition are presented for each class of node.

### 1.1.1 Server

Any node that makes access to certain services available to other nodes in the the network can be called a "server". In large networks, a dedicated server runs a special network operating system; in smaller installations, a non-dedicated server may run a personal operating system with peer-to-peer networking software running on top. A generic server typically has a more advanced processor, more memory, a larger cache, and more disk storage than a single-user workstation. A server may also have several processors rather than just one and may be dedicated to a specific support functions. Communications servers, modem servers, file servers, print servers, Web servers etc. are examples of different servers [10].

### 1.1.2 Client

Client node is the device or application that uses the services provided by a server. A client may be a PC or a workstation on a network using services provided from the network server, or it may be that part of an application program that runs on the workstation supported by additional software running on the server [10]. It is often the case that the clients request communication with another client, while the server manages certain statistics in the process.

### 1.1.3 Cluster

Clustering is a process of grouping servers and other network resources into a single system to elevate the network robustness in the event of failure of the resources. A network may have one or more clusters, depending on how big the network is. Clustering software adds a load-balancing feature to the clustering system, to make sure that processing is distributed in such as way as to optimize system throughput. In some signaling networks, clusters are groups of signaling points and individual signal transfer points.

### 1.1.4 Bridge

A bridge is a hardware device used to connect LANs so that they can exchange data. Bridges can work with networks that use different wiring or network protocols, joining two or more LAN segments to form what appears to be a single network. A bridge operates at the data-link layer of the Open Systems Interconnect reference model for computer-to-computer communications. It manages the flow of traffic between the two LANs by reading the address of every packet of data that it receives. In networks, where there are no bridges, certain elected nodes may create a communication network between neighbouring clusters and thus acts as a bridge. In such a network, a bridge is elected based on its visibility; it must be visible by all client stations of the clusters concerned and thus works as inter-cluster link.

## 1.2 Ad Hoc WLANs

The word "ad hoc" refers to making or happening only for a particular purpose or need, not planned in advance [42]. In networking context, ad hoc network is an IEEE 802.11 networking framework, in which nodes communicate directly with each other without the use of an access point, by which it can connect or communicate with the network. An ad hoc mode is also referred to as a peer-to-peer mode that is useful for establishing a network where infrastructure does not exist or where services are not required [14]. In ad hoc wireless LANs, all nodes work in ad hoc mode and form a network a dynamically without any existing infrastructure or topology. The nodes adjust accordingly with the topology change and hence are very robust. Also, since it does not utilize expensive network switches or other access and control points, it is a low cost solution. Due to the flexibility, robustness, and dynamic structure of such networks, ad hoc wireless LANs have made a way significantly into the business, military and personal communication sectors in a very short time [16]. The early groundbreaking research for ad hoc wireless LANs was supported by the Defense Advanced Research Projects Agency

(DARPA) and the Navy in US [41, 32]. Despite many advances over the last several decades in wireless communications, in general, and ad hoc wireless networks, in particular, the optimal design, scalability, performance, and fundamental capabilities of these networks remain poorly understood, at least in comparison with other wireless network paradigms and a lot of *"daylight"* remains in this field of research. However, with enormous potentials for such networks, ad hoc networks primarily support data networks, but it has been envisioned recently to enter to home networks, wireless device networks, distributed control systems, and sensor networks etc. In the following subsections, the network architecture, routing and scalability of an ad hoc network are described.

### 1.2.1 Network Architecture

As the name suggests, the most fundamental aspect of an ad hoc wireless network is that it does not have any pre-existing infrastructure. The challenge in design, topology and architecture of such networks stem from this characteristic. In comparison with conventional wireless networks, *viz.* cellular systems and wireless LANs, this kind of network offers extremely high flexibility. Unlike cellular systems, the nodes in this system have peer-to-peer communication between every two neighbouring nodes. Since there is no centralized control, in order to make effective communication successful among them, the nodes have to reconfigure themselves whenever the topology of the network changes. This is a dynamic process and is crucial to the system scalability and performance. The following are the reasons, why the topology of an ad-hoc network may change [16]:

1. *Node mobility*: Whenever the nodes are mobile, their positions may also change over time and topology may change.

2. *Change of power*: Power may suddenly change or fall off from certain communication node and which may result in different criteria for error-free reception mechanisms during a transmission process, resulting in topology change.

3. *Medium access control (MAC) algorithms*: Nodes that find access difficult through an existing topology and architecture may attempt it with a change in the topology.

4. *Flow dynamics*: Data flows come and go; so, if a node has nothing to transmit for sometime, its links are gone from the topology, to simplify the network further and improve scalability.

5. *Mode of nodes*: The mode of a node can either be sleeping or active; so, if a node goes to a sleeping mode, its links are gone from the topology, too.

Within the topology, certain nodes are close enough to be able to communicate with each other in a single hop. All nodes that can communication in single hop then forms a cluster that enables resource sharing among the nodes in a distributed manner and also to improve network reliability, scalability, and capacity [35, 4]. Clustered ad hoc WLANs can be operated with different modes of access systems; *viz.* bandwidth-on-demand (BoD) systems, Quality-on-Demand (QoD) systems etc. In this thesis we consider widely used bandwidth-on-demand (BoD) access mechanism. BoD is a dynamic system, where the access to the network resources is provided based on the bandwidth demand and defined by a set of rules by which nodes request transmission capacity from the network controller. The network controller is essentially an elected node within a cluster, which has the responsibility to share the requested bandwidth based on some fairness criterion and termed the "master node". This node decides on the allowed input and output rate based on the total cluster bandwidth, whenever any transmission request is generated within a cluster [31].

### 1.2.2 Routing

Routing algorithms decide certain feasible/optimal paths through which data transmission can take place. Before such path decisions can be taken, every node must have enough node and link statistics from its topology. Based on how the change of topology affects the routing decisions, networks can be either "combinatory stable" or "instantaneous". In combinatory stable networks, the change of topology is slow enough for the nodes to update link statistics to form a group. Ad hoc wireless local area networks (WLAN) are example of such networks. In instantaneous ad hoc networks, the topology changes take place very fast, links break and make very often and routing decisions becomes instantaneous and rather difficult. Such networks pose great challenges to research in dynamic routing decisions. Some wireless mobile ad hoc networks (MANET), which have been developed recently are examples of such network [25].

In some ad hoc networks, the nodes can adjust their power power accordingly to be able to transmit data in a single hop. The decision of adaptive power depends on certain perception of quality, *viz.* signal to noise and interference ratio (SINR), signal to noise ratio (SNR) etc. [11]. In such routing, nodes can send packets directly to their final destination via single hop routing as long as the link SINR is above a minimum threshold. However, the SINR is typically quite poor under single hop routing, and this method may also cause excessive interference to surrounding nodes. Also, despite its simplicity it is rather very expensive solution for large ad hoc networks. In large ad hoc wireless networks, packets are forwarded from source to destination through intermediate relay nodes. Since path loss causes an exponential decrease in received power as a function of distance, using intermediate relays can greatly reduce the total transmit power (the sum of transmit power at the source and all relays) needed for end-to-end packet transmission. Such routing is called as “multihop routing”. Essentially, in ad hoc networks, such routing is possible when some of the intermediate nodes act as bridges [21, 36]. Multihop routing using intermediate relay nodes is a key feature of ad hoc wireless networks: it allows for communication between geographically-dispersed nodes and facilitates the scalability and decentralized control of the network. However, it is much more challenging to support high data rates and low delays over a multihop wireless channel than over the single-hop wireless channels inherent to cellular systems and wireless LANs. This is one of the main difficulties in supporting applications with high data rate and low delay requirements, such as video, over an ad hoc wireless network [16].

### 1.2.3 Scalability

Scalability is a requirement for ad hoc wireless networks with a large number of nodes. It allows the complete ad hoc network to operate in an integrated manner. Due to large number of constraints and lack of centralized administration, scalability of ad hoc networks is still poorly understood [4]. The key to scalability lies in the use of distributed network control algorithms: algorithms that adjust local performance to account for local conditions. By forgoing the use of centralized information and control resources, protocols can scale as the network grows since they only rely on local information. Distributed protocols often consume a fair amount of energy in local processing and message exchange. Thus, trade offs arise between how much local processing should be done versus transmitting information to a centralized location for processing. This trade off is particularly apparent in sensor networks, where nodes close together have correlated data, and also coordinate in routing that data through the network. Many ad hoc network applications, especially sensor networks, could have hundreds to thousands of nodes or even more. The ability of existing wireless network protocols to scale to such large network sizes remains unclear [16].

### 1.2.4 Implementation Issues

An ad hoc wireless LAN network has certain advantages which make it an attractive business and personal solution and as a result such networks have made their ways into home networks, device networks, sensor networks and distributed networks within a very short period [16]. But certain implementation issues must be considered before choosing an ad hoc mode of operation in these networks [15]:

1. **Cost** : An ad hoc network leads to the ease of setting up a network without the need to purchase or install access points, which makes it financially a cheap and desirable option. But cost savings can easily be overrun by a bulk of complexities in bit rate performance if not properly implemented.

2. **Setup Time** : One of the basic advantages of ad hoc modes in wireless networks is that they are set up in a very quick time needing only to setup a network interface card for it to operate. But certain issues related to the channel properties and network size may take some calibration to be done before an ad hoc WLAN can be put into operation.

3. **Performance** : Issues related to performance must be well understood before any implementation is planned. In some small ad hoc networks, the network performance in terms of bit rate and QoS may be better than an administered one because no packet needs to travel through access points. However with large number of nodes, multiple access points to separate nodes onto non-overlapping channels to reduce medium access contention and collisions may reduce the system performance drastically. Also, because of a need for sleeping stations to wake up during each beacon interval, performance can be lower with an ad hoc mode due to additional packet transmissions if you implement power management.

4. **Limited Network Access** : Due to lack of a distribution system with ad hoc wireless LANs, nodes may not be allowed access to the Internet and other wired network services to a larger scale. In places, where there is a

strong need to access applications and servers on a wired network or Internet, an ad hoc WLAN may not be a suitable solution.

5. **Difficult network management** : Because of the fluidity of the network topology and the lack of a centralized device, the network management becomes much harder. The network performance, security audits etc. cannot be monitored because there is no defined access point in such networks. Effective network management with ad hoc wireless LANs requires this to take place at the user device level, which requires a significant amount of overhead packet transmission over the wireless LAN. This again steers ad hoc mode away from larger, enterprise wireless LAN applications.

## 1.3 Congestion Control

Congestion is an unwanted situation in networked systems, where the part of the network is being offered more traffic than its rated (desired) capacity. Congestion can be disastrous for a data transmission system as it manifests itself as depletion of resources that are critical to the operation of the system. These resources can be CPU, buffer space, bandwidth etc. Resource crunch will lead to lengthening of various queues for these resources. Due to finite length constraint, many packets may eventually get dropped, which, in turn, will deteriorate the response time of the system beyond permissible limits due to retransmission requests. *"Congestion control"* refers to the mechanism of combating congestion, which makes sure the resources are used optimally and the system has maximum data throughput with the given conditions.

The main objective of congestion control is to make sure the system is running at its rated capacity, even with the worst case overload situations. In certain systems, this is ensured by restricting certain nodes to transmit at the maximum capacity or to make use of certain resources monotonously. Doing this enables optimal usage of resources for all the nodes in the system with a measurable quality-of-service (QOS). In some systems, there are built-in mechanisms that does not allow congestion situation to take place and every node keeps track of system statistics and resources. This is often knows as *"congestion prevention"* or *"Congestion avoidance"*.

Congestion control is necessary for systems, whose nodes do not keep track of such statistics or do not keep resource information. In such systems, the nodes participate in the network, in which the topology changes very often and the network statistics also vary randomly. As such the control of congestion becomes an issue of the nodes that act as bridges. Ad hoc networks are examples of such scheme. In this thesis, the terms congestion control and congestion avoidance will be used synonymously with the ultimate aim to keep the total networked system free of congestion. Congestion control can either be rate-based control or buffer-based control depending on how the actual control is done. Most of the rate-based congestion control algorithms are applied during routing of data from node to node. In multihop routing, thus, congestion takes place on every hop and is termed hop-by-hop congestion control. However, for single hop routing congestion is only an end-to-end issue and more of rate adjustment of the source rather than destination. A major open challenge for research still remains for congestion control of large ad hoc wireless networks, where single hop routing is virtually impossible [36, 16].

## 1.4 Control System Concepts

In this thesis, a control-theoretic model is first developed for the system considered. The model is based on a time-delay model and designed according to the internal model control (IMC) principles. Also, to combat system instability, a Smith predictor (SP) is designed. In this section, relevant basic control system concepts are briefly presented.

### 1.4.1 Time delay Model

In process control, a time delay is the time it takes since the moment we make a change in the control input or signal until a reaction is seen in the output variable. The time-delay systems (called also hereditary or systems with after-effects) represent a class of infinite-dimensional systems largely used to describe propagation phenomena. Possible sources of time delays are: 1. The process may involve the transportation of materials or fluids over long distances. 2. The measuring device may be subject to long delays to provide a measurement. 3.The final control element may need some time to develop the actuating signal. Independently of the representation type, the effects of delay on the stability and control of dynamical systems (delays in the state and/or in the input) are problems of critical interest since the delay presence may induce complex behaviors (oscillations, instability, bad performance) for the closed-loop schemes: “small” delays may destabilize some systems, but “large” delays may stabilize others [27]. Indeed, for example, a sequence of delay ‘switches’ (stability to instability or instability to stability) may appear with the second order even for a single discrete or point delay in a linear differential-difference equation, if the delay value, seen as a parameter, is increased. Furthermore, a chaotic behavior may appear if the delayed state is a nonlinear function. But in other cases, chaotic systems may by stabilized by a delayed output. In control systems, it is well known that delays in feedback systems are accompanied by bandwidth ‘sensitivity’ to model uncertainty. Furthermore, delay perturbations due to some modeling errors may induce instability, and interconnection schemes of finite or infinite-dimensional systems with delay blocks may become unstable even if some “well-possessedness” property holds [17]. In this research, the control system model is built from a linear time-delay model. In Chapter 5, the effect of forward and backward time-delay is investigated.

### 1.4.2 Internal Model Control

The internal model control (IMC) is a control system result, which states that the control can be achieved only if the control system encapsulates, either implicitly or explicitly, some representation of the process to be controlled. If perfect control is to be achieved, the control scheme must be developed as an exact model based on IMC principles. In the open loop case when all the states of the particular process are known and the process is perfectly invertible. In practical, however, the process-model mismatch is common, which means the process may not be invertible and the system is often affected by unknown disturbances. In this case, IMC principles allow a closed-loop model to be implemented for achieving perfect or near-perfect model [18].

### 1.4.3 Smith’s Principle

In a process control, time-delay is crucial to system performance. The presence of time delays causes the following difficulties in process control:

1. A disturbance entering the process will not be detected until after a significant period of time.

2. The control action will be inadequate since its effects on a current error will affect the process variable only after a long delay.

3. Long time delays may originate instability in the system.

As such it is difficult to model a process that has time-delay, which often leads to unexpected results. One of the classical solutions to time-delays model compensation had been proposed in [37], known as a Smith predictor. The Smith predictor consists of an ordinary feedback loop plus an inner loop that introduces two extra terms directly into the feedback path. The first term is an estimate of what the process variable would look like in the absence of any disturbances. It is generated by running the controller output through a process model that intentionally ignores the effects of disturbances. The mathematical model used to generate the disturbance-free process variable has two elements connected in series. The first represents all of the process behaviour not attributable to dead time. The second represents nothing but the dead time. Subtracting the disturbance-free process variable from the actual process variable yields an estimate of the disturbances. By adding this difference to the predicted process variable, Smith created a feedback variable that includes the disturbances, but not the dead time. The Smith predictor essentially works to control the modified feedback variable (the predicted process variable with disturbances included) rather than the actual process variable. If it is successful in doing so, and if the process model does indeed match the process, then the controller will simultaneously drive the actual process variable towards the set point after set point changes or disturbances. Many of today's commercial PID controllers with time delay compensation use the Smith predictor strategy, or modifications from it. In this research, the Smith predictor is used to compensate for the loop delays to compensate for the forward disturbance in Sec. 3.1.

## 1.5 Problem Statement

Due to its distributed nature, flexibility, robustness and ease of installation, ad hoc wireless LAN has greatly increased the scope for research in wireless communications [16]. These LANs can be operated with different modes of access and routing systems; in this case, we consider multihop routing of data packets and bandwidth-on-demand (BoD) access mechanism.

In this thesis, a DSP-based solution is developed from control-theoretic paradigm to control the congestion of a multihop ad hoc wireless LAN with bandwidth-on-demand access. The design of the proposed system has been derived from the reference and error controller model, which essentially controls the queue length against a reference queue length based on the congestion that the system has to combat. This is done by regulating the desired input rate and required output rate such that the system is asymptotically stable in terms of all possible congestion scenarios. The Smith predictor is used in the closed-loop error controller to compensate for delays that could cause instability. Unlike conventional end-to-end feedback and stochastic control of congestion, this paper uses a hop-by-hop method. Hence, the control of congestion takes place on every hop to intermediate nodes that act as bridges. The obvious advantage of such control is fast reaction in each hop; however scalability remains a dilemma for such systems since flow adjustments are to be made on every hop. The network topology of the proposed system is considered as combinatory stable, which means that change in the network topology is slower than that required to update the network information by each node in the network. Underpinning the control paradigm, a filter based solution is proposed with an aim to ensure full-link utilization and to achieve maximum rate recovery as soon as the congestion has been cleared under system stability. This can then be used as a means of real-time control of congestion rather than on-demand control and mitigates the scalability problem to some extent. Filter models have been previously used in [40] to control the congestion of a ATM switching network, but here we also a novel congestion control solution for ad hoc wireless LANs. Simulation results are given to demonstrate the performance of the designed system. Similar congestion control algorithms had been developed by [30] and [31] but the proposed scheme improves on hardware solution, scalability and rate value limiting, to be illustrated later.

## 1.6 Overview of the Thesis

Based on the definitions presented in this chapter, the problem that this thesis deals with is defined. Also, brief definitions have been presented for certain terms that have been used throughout the thesis. In Chapter 2, research literatures that have contributed to the problem of congestion control over the years are briefly revised, as essential background of the present work. In this chapter, the classical congestion control in the Internet that had evolved due to different requirements have been illustrated in Sec. 2.1. Later, in Sec. 2.2, early background works for congestion control in the ad hoc wireless LAN are investigated. In Chapter 3, the control system model based on the IMC principle is derived. A basic system model have been presented for the derivation in Sec. 3.1 and reference and error controllers are designed from defined control objectives in Sec. 3.2. In Sec. 3.3, the designed model is analyzed and a time-delay stability analysis is done. Underpinning the control system model designed, a digital filter based solution is derived in Chapter 4. Also, the digital filter-based solution have been analytically investigated from stability and implementations perspective. Later, in Chapter 5, the digital filter-based solution have been simulated to test for system performance. The effects of time delays are investigated from different congestion scenarios in Sec. 5.2, while the backward, forward and combined congestion scenarios are simulated in Sec.s 5.1.1, 5.1.2 and 5.1.3, respectively. To assure the total system performance, MATLAB Simulink have been used for simulation, while for integrated performance in WLANs, an OPNET discrete event simulation tool (by MIL3) have been employed in this chapter. In Chapter 6, overall conclusions are drawn as directions for further research are given. Finally, the Simulink and OPNET simulation models are presented in the appendices.

# Chapter 2

# Literature Review

Congestion is an unwanted situation which takes place in the access points in the networks that have limited resources, such as buffer length. Also, the fact that large networks often have nodes having different input and output rates, congestion can take place in such access points, as well. Congestion control has been a serious issue in communication networks and is a key to network performance. Early works in congestion control are based on the modern Internet technology, where the access points are well defined and are administered by dedicated nodes. But due to lack of these infrastructures, congestion in ad hoc wireless LANs cannot be dealt in exactly the same way as that in the Internet, even though the basic purpose is the same. In the following sections, we look into different literatures to describe how congestion control algorithms have matured from that in Internet to ad hoc wireless LANs and how these algorithms affect the congestion in different scenarios.

## 2.1 Congestion in the Internet

Communication networks have experienced an explosive growth over the past decade, but the networks and resources have not grown up to same proportion. As such overwhelming growth of data have come under severe congestion problems. Much of the menace to packet loss in today's internet gateways lies in transport control protocol (TCP) implementations: the obvious ways to implement a window-based transport protocol can result in exactly the wrong behavior in response to network congestion. However, there had been series of congestion control and avoidance algorithms based on end-to-end or edge-to-edge flow control, unicast or multicast control. Most of these algorithms are rooted in the idea of achieving network stability by forcing the transport protocol to obey a packet conservation principle and by adjusting the transmission rate based on the loss probability [19].

In October of 1986, the Internet had the first of what became a series of congestion collapses. During this period, the data throughput from Laurence Berkeley National Laboratory (LBL) to University of California at Berkeley (UC, Berkeley) dropped from $32Kbps$ to $40bps$. Probed investigation into 4.3BSD (Berkeley Software Distribution) brought back the same answer; it was TCP at the root, which needed certain calibration at the congestion algorithm level. Following the dilemma, the following seven new algorithms came into existence in 4BSD [19, 26]:

1. Dynamic window sizing on congestion.
2. Round-trip-time variance estimation.
3. Exponential retransmit after back off.
4. Slow start.
5. More aggressive receiver acknowledge policy.
6. Karn's clamped retransmit back off.
7. Fast retransmit.

Algorithm 1 is based on resizing the transmission window size according to certain congestion scenario, like packet loss probability. To meet the needs of today's bandwidth-rich networks, Fisk and Feng developed dynamic right sizing algorithms of TCP packets in the event of congestion to allow much more fine tuned flow control for congestion control [12]. In algorithm 2, the transmission rate is adjusted by regulating the window size using round-trip-time (RTT) estimation. But if the round trip delay increases, the queue starts to form larger and situation may go out of control. To cope with such increased delays, Brakmo and Peterson in [6] proposed TCP Vegas, in which the retransmit rate was set proportional to the ratio of the RTT and the queuing delay [22]. RTT estimation has recently been renovated through adaptive Kalman filtering as proposed by Jacobsson *et al* in [20]. Congestion and queue management with congestion indication has also been attempted in various ways like random early detection techniques [13], random exponential marking techniques [3], which improved on algorithm 2. Algorithm 4 operates by observing that the rate at which new packets should be injected into the network is the rate at which the acknowledgments are returned by the other end. As a means of avoiding congestion rather than combating congestion scenarios, congestion avoidance algorithms also evolved. Basic Assumption of the congestion avoidance algorithm is that packet loss caused by damage is very small (much less than 1%), therefore the loss of a packet signals congestion somewhere in the network between the source and destination. There are two indications of packet loss: a timeout occurring and the receipt of duplicate ACKs. Congestion avoidance and slow start (algorithm 4) are independent algorithms with different objectives. But when congestion occurs, TCP must slow down its transmission rate of packets into the network, and then invoke slow start to get things going again. In practice they are implemented together [39]. Algorithm 7 evolved with certain changes to congestion avoidance algorithm in 1990 by Jacobson [19]. This was based on the assumption that TCP could retransmit as soon as the first acknowledgement (ACK) is seen, without having to wait for the repeated and delayed acknowledgements. This also depended on algorithm 5 for an aggressive acknowledgement for a faster retransmit attempt. After fast retransmit sends what appears to be the missing packet or segment, congestion avoidance, but not slow start can be performed. This was later also known as the fast recovery algorithm. It is an improvement that allows high throughput under moderate congestion, specially for large windows [39]. Algorithm 6 and algorithm 3 are basically same except for the fact that in the former the retransmission is clamped based on loss of packets after a backoff has happened, while in the later the retransmit attempts are made exponentially after a backoff, until successful transmission is done [19].

As far as use of control-theoretic model in the Internet is concerned, Mascolo first used a control-theoretic model for the flow-based congestion control of the traditional internet protocol in [24] as a time delay system. The author shows that the self-clocking principle, which is known to be a key component of any stable congestion Internet control algorithm, corresponds to simple proportional controller plus a Smith predictor (SP), which overcomes feedback delays that are due to propagation times.

## 2.2 Congestion in Ad Hoc WLANs

In recent years, depending on how challenging the congestion problem is, a combination of the algorithms in Sec. 2.1 work satisfactorily under congestion circumstances [19]. As issue related to mobility of computing and business came stronger in the recent years, ad hoc WLANs became a popular choice with added algorithmic challenges. Owing to its high potential, ad hoc networks became a standard in IEEE 802.11 networking framework [14, 16]. Compared to internet technology, ad hoc WLANs do not have any centralized or distributed control points. Thus, when an ad hoc WLAN becomes moderately large, it becomes increasingly more difficult to maintain high data rate with BoD systems using the existing algorithms. Of many reasons as to why the existing algorithms do not fit in the ad hoc network framework, are issues related to the change of topology and increased complexity for system without administration and without access point.

Similar to the ones illustrated in Sec. 2.1, a congestion control scheme was proposed based on end-to-end feedback exchanges in [38]. But soon it was found out that such TCP based solution do not work well for bandwidth-on-demand ad hoc WLAN systems due to lack of control access points. Later, as one of the early works in congestion control for ad hoc networks, Altman *et al* in [2] showed that the congestion control problem can be formulated as a stochastic control problem. The paper considers the design of explicit rate-based congestion control for high-speed communication networks and shows that this can be formulated as a stochastic control problem where the controls of different users enter the system dynamics with different delays. It also shows the existence, derivation and the structure of the optimal controller, as well as of suboptimal controllers of the certainty-equivalent type with defined context for congestion control. In particular, this paper considers two certainty-equivalent controllers which are easy to implement, and show that they lead to bounded infinite-horizon average cost, and stable queue dynamics with simulations. However, certain users may suffer excessive delays beyond limits. All these methods demanded use of continual statistical information and complexity in order to have a proper control on congestion.

In [23], for high-speed communication networks, which are characterized by large bandwidth-delay products, the adverse impact on the stability of closed-loop congestion control algorithms is found out. To combat these effects the classical control theory model and Smith's principle are proposed as key tools for designing an effective and simple congestion control law for high-speed data networks. The authors make argument through mathematical analysis to show that the proposed control law guarantees stability of network queues and full utilization of network links in a general network topology and traffic scenario during both transient and steady-state condition. Also, the authors makes a comparison with the explicit rate indication for congestion avoidance (ERICA) algorithm with necessary transformation for the control law to a window, to be applied in the Internet.

In [33], an H-infinity controller is designed guaranteeing stability robustness with respect to uncertain time-varying multiple delays. It also brings the queue length at the bottleneck node to the desired steady-state value asymptotically and satisfies a weighted fairness condition. Lower bounds for stability margins for uncertainty in the time-delays and for the rate of change of the time-delays are derived and time-domain performance of the controller is demonstrated by a number of simulations.

Both of these congestion control scheme opened the wide range of scope to combat against congestion in classical control-theoretic method. As far as early ad hoc WLANs with BoD protocols are concerned, Açar and Rosenburg successfully used such protocols for satellite networks in [1], where the problem is formulated as a optimization problem. The authors in this paper present a demand assignment multiple access based resource management protocol, weighted fair bandwidth-on-demand (WFBoD), for geostationary satellite networks with on-board processing, which combines flexibility with efficiency and the right level of traffic segregation. The paper tries to formulate the global resource allocation problem, central to which is a large generic integer value optimization problem with a large number of coupled constraints and proposes heuristics to solve this optimization problem and compare their performances and their complexity with the formal solution.

Priscoli and Pietrabissa, in [30], used BoD protocol along with Smith's predictor to ensure that the queue lengths are controlled by reference values for a satellite terminal, while in [31], authors models the problem of congestion control in geostationary satellite framework with large delays in the form of two cascade controllers: on-board and on-earth. It uses similar time-delay based control-theoretic model to propose the solution. In [31], similar protocol and algorithm has been used in wireless LANs by the same authors. In these papers, the authors present a model based control methodology to simultaneously computer the capacity requests necessary to access the network and the capacity allocations required to regulate the rates of the traffic flows. The scheme proposed by the authors also allows to compute upper bounds of the queue lengths in all network buffers. The high speed wireless LAN considered in [31] has been developed within the European Union project wireless indoor flexible high bit rate modern architecture (WINDFLEX). In [9], the BoD protocol uses a adaptive predictor coupled with a receding horizon controller.

Congestion control in multihop wireless networks is a hop-by-hop process assuming that the intermediate nodes help in the routing acting as bridges. Yi and Shakkottai investigated hop-by-hop congestion control algorithm in [43]. The authors focus on congestion control over multihop, wireless networks using hop-by-hop congestion control. Also in their work, time-division strategy for medium access control algorithm has been used for channel access, such that at any point in space, the physical channel can be accessed by a single user at each instant of time. A fair hop-by-hop congestion control algorithm with the MAC constraint being imposed in the form of a channel access time constraint is developed, using an optimization based framework. The authors also show that the algorithm is globally stable using a Lyapunov function based approach and shows that the hop-by-hop control algorithm has the property of spatial spreading. For simulation, bounds on the peak load at a node are also derived, both with hop-by-hop control, as well as with end-to-end control.

Congestion control and scheduling algorithms for wireless networks has been shown in an integrated manner by Priscoli and Isidori in [29], which deals with the problem of guaranteeing a target quality of service (QoS) to connections set-up over wireless internet protocol (IP) networks, while efficiently exploiting the air interface. This problem is then coped with congestion control and traffic scheduling algorithms: congestion control deals with the problem of computing the traffic relevant to in progress connections which can be admitted into the wireless network without causing the infringing of the QoS, while the scheduling deals with the problem of deciding the priorities for the transmission of the admitted traffic over the air interface. The authors also present the original and simple architecture and procedures of a traffic control module aiming at solving the problem following a control-based approach. The controller steers the overall system towards an ideal equilibrium at which desirable performance is achieved and is in charge of periodically updating this ideal equilibrium, which is a function of the IP traffic presently offered to the wireless network.

Digital filters are key technology to todays fast communications and signal processing. Digital filter-based approach to congestion control has been shown by Tan *et al.* The authors in [40] proposes a control-theoretic approach to design rate-based controllers in order to flow-regulate the best-effort service in asynchronous transfer mode (ATM) switching networks. The proposed control by the authors uses a recursive digital filtering controller as a original approach, where the control parameters can be designed to ensure the stability of the control loop in a control-theoretic sense. The stability of closed-loop congestion control system is analyzed by SchurCohn stability criterion, which leads to certain necessary and sufficient stability condition under which the controlled ATM switching network is asymptotically stable in terms of buffer occupancy and also to ensure a fair share of the available bandwidth at the bottleneck node can be achieved according to the proposed control policy.

Most of the congestion control schemes described in this chapter are either end-to-end feedback or stochastic control scheme. However, this research exploits a hop-by-hop method for congestion control of ad hoc wireless LANs using digital filters. The control of congestion for this scheme takes place on every hop to intermediate nodes that act as bridges. The obvious advantage of such control is fast reaction in each hop with scalability as a dilemma since flow adjustments are to be made on every hop. The network topology of the proposed system is considered as combinatory stable, which means that change in the network topology is slower than that required to update the network information by each node in the network. Underpinning the control theoretic model that will be developed in Chapter 3, a discretized and properly approximated model is then developed in chapter 4 towards a digital-filter based solution for control of congestion. The proposed system can be used as a real-time control of congestion rather than an on-demand control. Since the solution is hardware based, it also mitigates the scalability problem to some extent. Unlike [40], the proposed scheme introduces a novel congestion control solution for ad hoc wireless LANs. Simulation results are given to demonstrate the performance of the designed system later in Chapter 5.

# Chapter 3

# Control System Design

In this chapter, a control-system model is formulated according to internal model control (IMC) principle. The internal model control objectives are derived from congestion avoidance perspective with congestive disturbances entering the system. The purpose of the controller is to ensure that the system regulates the input and output rates based on the bandwidth demands and disturbances. Also, in this chapter time-delay stability analysis will be carried out to demonstrate the system performance.

## 3.1 System Model

In order to derive the system model, a basic ad hoc wireless network with BoD access is studied first. As shown in Figure 3.1, in each cluster the master nodes are chosen by the nodes, which has the responsibility to share the cluster capacity among the nodes. Also, cluster nodes choose one of them to act as the bridge, which is visible to the neighbouring clusters. In Figure 3.1 nodes $B$, $E$ and $I$ are shown as master nodes and nodes $D$ and $H$ act as bridges at a particular instant of time. The choices of master nodes and the bridges may change as the topology changes. In each cluster, every node can communicate with all others in single hop. To simplify the network diagram, only master nodes are shown with end-to-end wireless links.

We consider a typical case when node $A$ of cluster 0 wants to transmit to node $M$ of cluster 2. The step by step procedures involved are described next.

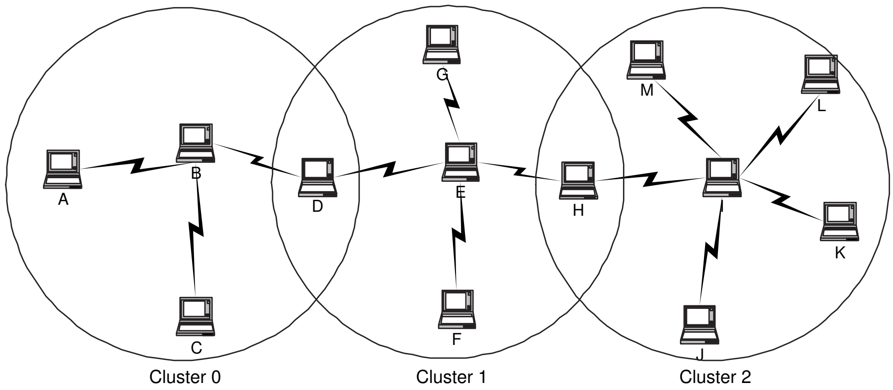


FIGURE 3.1: A typical ad hoc wireless LAN

1. In its turn, in the time-division cycle, node $A$ requests its share of bandwidth from master node $B$ of the cluster 0. Node $D$ will also announce to the master node $B$, the maximum desired input rate, $r_{\mathrm{DES}}(t)$, at which it is able to accept. The source controller in node $B$ decides at what rate it should output the data.

2. Once node $B$ has complete knowledge of which node wants what share, it then announces what bandwidth can be given to each node in the cluster depending on the available bandwidth, bandwidth requests from nodes in the cluster and a fairness criterion. Master node $B$ also announces back to bridge node $D$ what rate, $r_{\mathrm{IN}}(t)$, it should be allowed to take as input. The delay that occurs between the desired input rate announcement and the acknowledged maximum input rate between bridge $D$ and master node $B$, is called the backward delay and will be denoted by $T_{\mathrm{BW}}$ here. If the desired input rate, $r_{\mathrm{DES}}(t)$, is higher than the acknowledged input rate, $r_{\mathrm{IN}}(t)$, an additive backward disturbance takes place, which is denoted by $d_{\mathrm{BW}}(t)$. The controller must decide on the rates based on the backward disturbance created from differences from the master node.

3. Bridge $D$ receives data from node $A$ in a single hop, with associated input rate of $r_{\mathrm{IN}}(t)$, derived from the source and bridge controller and disturbance inputs. Node $D$ also announces to master node of cluster 1, node $E$, at what output rate, $r_{\mathrm{REQ}}(t)$, it wants to transmit. Node $H$, connecting the clusters 1 and 2, also announces what rate it is able to receive as $r_{\mathrm{DES}}(t)$.

4. Node $E$ feeds the output rate, $r_{\mathrm{OUT}}(t)$, back to node $D$, at which it is allowed to transmit. The time delay that occurs between the announced rate and the maximum allowed output rate is called the forward loop delay and will be denoted by $T_{\mathrm{FW}}$. If the requested output rate, $r_{\mathrm{REQ}}(t)$, is higher than the

allowed output rate, $r_{\text{OUT}}(t)$, an additive forward disturbance takes place, which is denoted by $d_{\text{FW}}(t)$. This disturbance is critical to system stability. The controller must decide on the rates based on the backward disturbance created from differences from master node.

5. Node $D$ now transmits data to node $H$, at a rate $r_{\text{OUT}}(t)$ allowed by the bridge controller, which is decided based on the disturbances and maximum throughput rate. Node $H$ requests the required output rate, $r_{\text{REQ}}(t)$, at which it wants to transmit.

6. Master node $I$ allows the rate of $r_{\text{OUT}}(t)$ as in step 4. At this moment, node $H$ dispatches the data directly to the node $M$ of cluster 2 at a rate of $r_{\text{OUT}}(t)$. The destination controller decides at which rate it can receive data, to avoid overflow of queues and buffers.

In the procedures described above, the source, destination and intermediate nodes incrementally start to form queues when the actual input rate, $r_{\text{IN}}(t)$, is higher than the actual output rate, $r_{\text{OUT}}(t)$, for them. The proposed filter based solution therefore makes sure that the decisions of allocating input and output rate are optimally controlled in the source, destination and in the bridge controller. In order to arrive at such solution a time delay model based control system will first be derived. Later, discretization will be done assuming that the forward and backward delays are integer multiples of the unit sample time, $T_s$. The present control system will have two distinct parts: the reference controller and error controller. A reference queue length, $q_{\text{REF}}(t)$, will be first deduced from the maximum data rate, $r_{\text{MAX}}(t)$, and backward disturbance, $d_{\text{BW}}(t)$, in the reference controller. The closed-loop error controller will control and minimize the error between instantaneous queue length, $q(t)$, and reference queue length, $q_{\text{REF}}(t)$. The overall system is intended to meet the following **objectives**:

1. The queue lengths must be limited to maximum buffer size of $S$, with a lower limit of 0, i.e.
$$0 \leq q(t) \leq S \quad . \tag{3.1}$$
This will make sure that no data is lost due to overflow and that link utilization is maximum.

2. When congestion is cleared by the congestion controller, the controller must also make sure the desired input and required output rates, $r_{\text{DES}}(t)$ and $r_{\text{REQ}}(t)$ respectively, are driven back to the maximum data rate, $r_{\text{MAX}}(t)$,

as if congestion was not there. This must be immediate to ensure that the system throughput is maximum at all states.

3. The system impulse response must be asymptotically stable independently of forward or backward disturbance.

## 3.2 Design of the Controllers

To meet the control missions, it is necessary to model the system in line with the control objectives set. In the following sections, a reference and error controller model for the bridge controller will be derived in the Laplace domain in accordance with [31] and later a digital filter-based solution is presented with necessary modifications. While both the controllers are based on a time-delay model, the former controller is based on open loop IMC principle and the later is based on closed-loop IMC principle.

### 3.2.1 Reference Controller Model

As mentioned before, the queue length, $q(s)$, at a bridge controller node will incrementally build up when the input rate, $r_{\text{IN}}(s)$, is higher than the output rate, $r_{\text{OUT}}(s)$, i.e.

$$q(s) = \frac{1}{s}\left[r_{\text{IN}}(s) - r_{\text{OUT}}(s)\right] \quad . \tag{3.2}$$

In order to make sure that the objective 1 is met, the necessary condition is $\frac{1}{s}(r_{\text{IN}}(s) - r_{\text{OUT}}(s)) \leq S$, where $S$ is the maximum buffer length. In [34] the buffers are considered unrealistically large to contain all the receive traffic, while in this paper the minimum buffer size to allow controllability will be defined realistically later in equation (3.18).

We have previously defined that, forward and backward disturbances are given by

$$\begin{aligned} d_{\text{FW}}(s) &= r_{\text{REQ}}(s) - r_{\text{OUT}}(s) \quad , & (3.3) \\ d_{\text{BW}}(s) &= r_{\text{DES}}(s) - r_{\text{IN}}(s) \quad , & (3.4) \end{aligned}$$

where $r_{\mathrm{DES}}(s)$ and $r_{\mathrm{REQ}}(s)$ are the desired input rate and required output rate respectively and $r_{\mathrm{OUT}}(s)$ and $r_{\mathrm{IN}}(s)$ are the output and input rates allowed by the master node. It should also be noted that, since all rates are bound by the maximum throughput rate of $r_{\mathrm{MAX}}(s)$ and since $r_{\mathrm{OUT}}(s) \leq r_{\mathrm{REQ}}(s)$ and $r_{\mathrm{IN}}(s) \leq r_{\mathrm{DES}}(s)$, values of forward and backward disturbances are bound by

$$0 \leq d_{\mathrm{FW}}(s) \leq r_{\mathrm{REQ}}(s) \leq r_{\mathrm{MAX}}(s) \quad , \tag{3.5}$$

$$0 \leq d_{\mathrm{BW}}(s) \leq r_{\mathrm{DES}}(s) \leq r_{\mathrm{MAX}}(s) \quad . \tag{3.6}$$

With the disturbances present, to establish a reference queue length, $q_{\mathrm{REF}}(s)$, the reference controller must define the reference desired input rate, $r_{\mathrm{DES}_{\mathrm{REF}}}(s)$ and reference required output rate, $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ based on the maximum throughput rate. Growth of $q(s)$ is then controlled against $q_{\mathrm{REF}}(s)$ through control over input and output rates as shown in equation (3.2). To achieve such control, reference rates and queue length must be defined.

In order to satisfy the congestion control objective 2, the reference desired rate should always be set the maximum throughput rate,

$$r_{\mathrm{DES}_{\mathrm{REF}}}(s) = r_{\mathrm{MAX}}(s) \quad . \tag{3.7}$$

From equations (3.5) and (3.6), it is evident that the maximum required output reference rate, $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$, can be $r_{\mathrm{MAX}}(s)$, and this happens when there is no backward disturbance is present. Also, since maximum backward disturbance is also bound by equation (3.6) as $r_{\mathrm{MAX}}(s)$, the minimum value of $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ is 0. Thus, with the given constraint in equation (3.6), $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ is given by

$$r_{\mathrm{REQ}_{\mathrm{REF}}}(s) = r_{\mathrm{MAX}}(s)e^{-sT_{\mathrm{BW}}} - d_{\mathrm{BW}}(s) \tag{3.8}$$

where the delay term, $e^{-sT_{\mathrm{BW}}}$, is due to backward delay, $T_{\mathrm{BW}}$, in the time-delay model caused by the master node in the backward acknowledgment process. Equation (3.8) conforms with the internal model principle, stating that the closed-loop system behaves much more like an open loop system in the absence of the disturbances [5].

Equations (3.7) and (3.8) show that the required reference output rate, $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$, can be redefined in terms of the desired reference output rate, $r_{\mathrm{DES}_{\mathrm{REF}}}(s)$,

$$r_{\mathrm{REQ}_{\mathrm{REF}}}(s) = r_{\mathrm{DES}}(s)e^{-sT_{\mathrm{BW}}} - d_{\mathrm{BW}}(s) \quad . \tag{3.9}$$

During the time interval between requesting an output rate and acknowledgment of the permitted rate, packets will integrally get accumulated at the buffer. A reference value of the queue length in the reference controller can thus be modeled by

$$q_{\mathrm{REF}}(s) = \frac{1}{s}\left[r_{\mathrm{DES}_{\mathrm{REF}}}(s) - r_{\mathrm{REQ}_{\mathrm{REF}}}(s)e^{-sT_{\mathrm{FW}}}\right]. \tag{3.10}$$

Combining equations (3.7), (3.8) and (3.10), $q_{\mathrm{REF}}(s)$ can further be expressed in terms of inputs $r_{\mathrm{MAX}}(s)$ and $d_{\mathrm{BW}}(s)$ as

$$q_{\mathrm{REF}}(s) = \frac{1}{s}[r_{\mathrm{MAX}}(s)(1 - e^{-s(T_{\mathrm{BW}}+T_{\mathrm{FW}})}) + d_{\mathrm{BW}}(s)e^{-sT_{\mathrm{FW}}}] \quad . \tag{3.11}$$

The delays in equation (3.11) are due to the time-delay model used in the controller. Knowledge of bounded values of time delays are crucial to stability of time-delay systems [17]. In this system we have a time-division multiplexed demand assignment cycle at the master node and these delays are explicitly known. This is essentially the requirement for the control mission set in objective 3.

### 3.2.2 Error Controller Model

Based on the disturbances and the reference rates, the error controller must control the values of $r_{\mathrm{IN}}(s)$ and $r_{\mathrm{OUT}}(s)$ in closed-loop fashion in order to minimize the error, $e(s)$, between $q(s)$ and $q_{\mathrm{REF}}(s)$. Controlling the values of $r_{\mathrm{DES}}(s)$ and $r_{\mathrm{REQ}}(s)$ would directly affect these rates, if a time-delay model is used, as in Sec. (3.2.1). The control would stem from the effect of worst case congestion, $d_{\mathrm{FW}}(s)$, which has been neglected in the reference controller.

As shown in equations (3.3) and (3.4), to compensate for the effects of $d_{\mathrm{FW}}(s)$, it is necessary to reduce $r_{\mathrm{DES}}(s)$ rather than to reduce $r_{\mathrm{REQ}}(s)$. This enables a fine-tuned control for the output rate such that congestion can be avoided and controlled. This control is done based on the feedback error between the reference queue length, $q_{\mathrm{REF}}(s)$ and the instantaneous queue length, $q(s)$, given by

$$e(s) = q(s) - q_{\text{REF}}(s) \quad . \tag{3.12}$$

Since the forward disturbance happens after $r_{\text{OUT}}(s)$ has been announced by the master node after $T_{\text{FW}}$, the compensation of $d_{\text{FW}}$ during this time requires that the error controller waits for the "dead time", $T_{\text{BW}}$, for the next error feedback to be available. This can cause instability and as such, control of $r_{\text{IN}}(s)$ and $r_{\text{OUT}}(s)$ by means of $e(s)$ must consider the effect of "dead time" according to Smith's principle [37]. Hence, transfer function between $e(s)$ and $d_{\text{FW}}(s)$ can be given by

$$\frac{e(s)}{d_{\text{FW}}(s)} = \frac{1 - e^{-sT_{\text{BW}}}}{s} + \frac{e^{-sT_{\text{BW}}}}{s + \frac{1}{T_K}}. \tag{3.13}$$

$T_K$ is the time constant, defined as the difference between minimum buffer-filling time and total delays, given by

$$T_K = \frac{S}{r_{\text{MAX}}} - (T_{\text{FW}} + T_{\text{BW}}) \quad . \tag{3.14}$$

Since both terms in the sum in equation (3.13) are first order systems and have poles at $s = 0$ and at $s = -\frac{1}{T_K}$, the system is stable as a requirement from control objective 3. Equation (3.13) also establishes the limiting values of $e(s)$. Taking the inverse Laplace transform of equation (3.13) yields

$$e(t) = \int_{\text{t}-\text{T}_{\text{BW}}}^{t} d_{\text{FW}}(\tau)d\tau + \int_{0}^{t-T_{\text{BW}}} e^{-\frac{\tau}{\text{T}_\text{K}}} d_{\text{FW}}(t - T_{\text{BW}} - \tau)d\tau. \tag{3.15}$$

Considering $T_K > 0$, $d_{\text{FW}}(t) = 0$ for $-(T_{\text{BW}} + T_{\text{FW}}) \leq t \leq 0$ and $0 \leq d_{\text{FW}}(t) \leq r_{\text{MAX}}$ for $t >= 0$ in equation (3.15) yields

$$\begin{aligned} 0 \leq e(t) \quad &\leq \quad r_{\text{MAX}} \left\{ \int_{\text{t}-\text{T}_{\text{BW}}}^{t} d\tau + \int_{0}^{t-T_{\text{BW}}} e^{-\frac{\tau}{T_K}} d\tau \right\} \\ &\leq \quad r_{\text{MAX}} \left( T_{\text{BW}} + T_K \right), \forall t \geq 0 \quad . \end{aligned} \tag{3.16}$$

Equation (3.16) is essentially limiting the queue error, $e(s)$, that is set by the controller and can be verified. In equation (3.9), when $d_{\text{BW}}(s) = 0$, $r_{\text{REQ}_{\text{REF}}}(s)$

is asymptotically driven to the value of $r_{\mathrm{DES}_{\mathrm{REF}}}(s)$ and $q_{\mathrm{REF}}(s) = 0$. But when $d_{\mathrm{BW}}(s) \neq 0$, $q_{\mathrm{REF}}(s) > 0$ according to equation (3.10). Thus, the minimum error takes place when $q(s)$ and $q_{\mathrm{REF}}(s)$ are same, i.e.

$$e(s)_{\mathrm{MIN}} = 0 \quad . \tag{3.17}$$

On the other hand, from equation (3.12), it is evident that $e(s)$ is maximum when $q(s)$ is maximum and $q_{\mathrm{REF}}$ is minimum. According to control objective 1 and equation (3.14), the upper limit of $q(s)$ is given by

$$\begin{aligned} q_{\mathrm{MAX}}(s) &= r_{\mathrm{MAX}}(s)\left[e^{-sT_K} + e^{-sT_{\mathrm{BW}}} + e^{-sT_{\mathrm{FW}}}\right] \\ &= S \quad . \end{aligned} \tag{3.18}$$

From equation (3.10), $q_{\mathrm{REF}}(s)$ is minimum when $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ is maximum as shown in the inequality (3.5). The maximum value of $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ from equation (3.8) is $r_{\mathrm{MAX}}(s)$. Thus, $q_{\mathrm{REF}_{\mathrm{MIN}}}(s)$ and the maximum error, $e_{\mathrm{MAX}}(s)$ can be shown as

$$q_{\mathrm{REF}_{\mathrm{MIN}}}(s) = \frac{1}{s} r_{\mathrm{MAX}}(s)[1 - e^{-sT_{\mathrm{FW}}}] = r_{\mathrm{MAX}}(s)e^{-sT_{\mathrm{FW}}} \quad , \tag{3.19}$$

$$e_{\mathrm{MAX}}(s) = q_{\mathrm{MAX}}(s) - q_{\mathrm{REF}_{\mathrm{MIN}}}(s) = r_{\mathrm{MAX}}(s)\left[e^{-sT_K} + e^{-sT_{\mathrm{BW}}}\right] \quad . \tag{3.20}$$

In the time-domain, equations (3.17) and (3.20) can be combined as the following inequality

$$0 \leq e(t) \leq r_{\mathrm{MAX}}(T_K + T_{\mathrm{BW}}) \quad . \tag{3.21}$$

Inequality (3.21) is a direct result of control objectives 1 and 3. With known limits set in inequality (3.21), the controller must be able to control $q(s)$ by controlling the rates $r_{\mathrm{DES}}(s)$ and $r_{\mathrm{REQ}}(s)$. This would indirectly affect actual input and output rates, $r_{\mathrm{IN}}(s)$ and $r_{\mathrm{OUT}}(s)$ from the master node, resulting from an additive forward disturbance of $d_{\mathrm{FW}}(s)$. The control of $r_{\mathrm{DES}}(s)$ and $r_{\mathrm{REQ}}(s)$ can be done using a error compensation method through desired input rate error, $r_{\mathrm{DES}_{\mathrm{ERR}}}(s)$ and required input rate error, $r_{\mathrm{REQ}_{\mathrm{ERR}}}(s)$ as

$$r_{\mathrm{DES}}(s) = r_{\mathrm{DES_{REF}}}(s) - r_{\mathrm{DES_{ERR}}}(s), \tag{3.22}$$

$$r_{\mathrm{REQ}}(s) = r_{\mathrm{REQ_{REF}}}(s) - r_{\mathrm{REQ_{ERR}}}(s). \tag{3.23}$$

As specified at the beginning of Sec. 3.2.2, this compensation will be in favour of $r_{\mathrm{DES}}(s)$ and not of $r_{\mathrm{REQ}}(s)$ due to advantages in fine-tuned control. Using the dead time compensation method for the $T_{\mathrm{BW}}$ by Smith's principle [37], we have the following transfer functions for the error rates,

$$\frac{r_{\mathrm{DES_{ERR}}}(s)}{e(s)} = \frac{s\frac{1}{T_K}}{s + \frac{1}{T_K}(1 - e^{-sT_{\mathrm{BW}}})}, \tag{3.24}$$

$$\frac{r_{\mathrm{REQ_{ERR}}}(s)}{e(s)} = 0. \tag{3.25}$$

Multiplying the transfer functions in equations (3.24) and (3.13), the following Laplace and time domain transfer functions can be shown,

$$\frac{r_{\mathrm{DES_{ERR}}}(s)}{d_{\mathrm{FW}}(s)} = \frac{\frac{1}{T_K}}{s + \frac{1}{T_K}}, \tag{3.26}$$

$$r_{\mathrm{DES_{ERR}}}(t) = \frac{1}{T_K}\int_0^t e^{-\frac{\tau}{T_K}} d_{\mathrm{FW}}(t-\tau)d\tau. \tag{3.27}$$

From equation (3.27) it follows that when $d_{\mathrm{FW}}(t) = 0$,

$$r_{\mathrm{DES_{ERR}}}(t) = 0 \quad ,$$

$$r_{\mathrm{DES}}(t) = r_{\mathrm{MAX}}(t) \quad .$$

This is the necessary target to be achieved in control objective 2. From Equation (3.26) it is also evident that the control system between $r_{\mathrm{DES_{ERR}}}(s)$ and $d_{\mathrm{FW}}(s)$ is a first order stable system with a pole at $s = -\frac{1}{T_K}$ [control objective 3]. When $0 \leq d_{\mathrm{BW}}(t) \leq r_{\mathrm{MAX}}(t)$, The compensation through $r_{\mathrm{DES_{ERR}}}(t)$ decreases exponentially with time constant, $T_K$ and as such, the following inequality holds

$$0 \leq r_{\mathrm{DES_{ERR}}}(t) \leq r_{\mathrm{MAX}}(t) \quad . \tag{3.28}$$

On the other hand, since $r_{\mathrm{REQ_{ERR}}}(t) = 0$ according to equation (3.25), the controller sets the desired input rate to be low enough to allow the output rate be higher than input rate when a congestion is to be cleared and high enough to allow the input rate be higher than the output rate when the effects of forward disturbance are to be mitigated. This would mean that the required output rate, $r_{\mathrm{REQ}}(s)$, is bound by equation (3.5) as

$$0 \leq r_{\mathrm{REQ}}(s) \leq r_{\mathrm{MAX}}(s) \quad . \tag{3.29}$$

Since the control system is based on a time-delay model, the time delays play an important role in the system performance. The larger the time delays are, the slower and more unstable the control action becomes and the harder it becomes to control the congestion with the time-delay model [8].

## 3.3 Performance Analysis

The open loop reference controller derived in Sec. 3.2.1 and the closed-loop error controller derived in Sec. 3.2.2 are the two time-delay systems based on IMC principles that controls the congestion in integrated manner. However, certain analysis of the proposed controller needs to be done in order to investigate the performance in the presence of congestive disturbances. In order to analyze the performance, we present the following lemmas and theorems and their proofs and later we also analyze the time-delay performances that may affect the system.

### 3.3.1 Theoretical Performance

**Lemma 3.1.** *When initially $q_{\mathrm{REF}}(t) = 0$, $d_{\mathrm{BW}}(t) = 0$ for $[-T_{FW} \leq t \leq 0]$ and $0 \leq d_{\mathrm{BW}}(t) \leq r_{\mathrm{MAX}}(t), \forall t \geq 0$, the reference controller derived equation (3.11) has the following properties:*

1. *$0 \leq q_{\mathrm{REF}}(t) \leq r_{\mathrm{MAX}}(t)T_{\mathrm{FW}}, \forall t \geq 0$, which eventually means full-link utilization and overflow avoidance.*

2. *$r_{\mathrm{DES_{REF}}}(t) = r_{\mathrm{REQ_{REF}}}(t) = r_{\mathrm{MAX}}(t), \forall t \geq t_{\mathrm{BC}} \geq T_{\mathrm{BW}}$, when $d_{\mathrm{BW}}(t) = 0$ for $t \geq t_{\mathrm{BC}}$.*

**Proof:** From equations (3.7) and (3.8), we have the following definition of the desired input and required output rates and also the reference queue length:

$$r_{\mathrm{DES_{REF}}}(s) = r_{\mathrm{MAX}}(s) \quad ,$$

$$r_{\mathrm{REQ_{REF}}}(s) = r_{\mathrm{MAX}}(s)e^{-sT_{\mathrm{BW}}} - d_{\mathrm{BW}}(s) \quad ,$$

$$q_{\mathrm{REF}}(s) = \frac{1}{s}[r_{\mathrm{MAX}}(s)(1-e^{-s(T_{\mathrm{BW}}+T_{\mathrm{FW}})}) + d_{\mathrm{BW}}(s)e^{-sT_{\mathrm{FW}}}] \quad .$$

By computing the inverse Laplace transform of equation (3.11), we have

$$q_{\mathrm{REF}}(t) = \int_{t-T_{\mathrm{FW}}}^{t} [r_{\mathrm{MAX}}(\tau - T_{\mathrm{BW}}) - d_{\mathrm{BW}}(\tau)]d\tau \quad . \tag{3.30}$$

Since $r_{\mathrm{MAX}}(t)$ is generally step-like function and constant and also since $d_{\mathrm{BW}}(t) = 0$ for $[-T_{FW} \leq t \leq 0]$ and $0 \leq d_{\mathrm{BW}}(t) \leq r_{\mathrm{MAX}}(t), \forall t \geq 0$, property 1 is satisfied. Also since $r_{\mathrm{MAX}}(t) = r_{\mathrm{MAX}}(t - T_{\mathrm{BW}})$, from equation (3.8) and equation (3.7), it can be shown that, when $d_{\mathrm{BW}}(t) = 0$, $r_{\mathrm{DES_{REF}}}(t) = r_{\mathrm{REQ_{REF}}}(t) = r_{\mathrm{MAX}}(t)$. This is the property 2.

**Lemma 3.2.** *Initially when $e(t) = 0, T_K > 0, d_{\mathrm{FW}}(t) = 0$ for $[-T_{\mathrm{BW}} \leq t \leq 0]$ and $0 \leq d_{\mathrm{FW}}(t) \leq r_{\mathrm{MAX}}(t), \forall t \geq 0$, the error controller derived equation (3.24) and equation (3.25) has the following properties:*

1. $0 \leq e(t) \leq r_{\mathrm{MAX}}(t)[T_{\mathrm{BW}} + T_K], \forall t \geq 0$, *which essentially sets the error limits.*
2. $r_{\mathrm{DES_{ERR}}}(t) = 0, \forall t \geq t_{\mathrm{FC}} \geq 0$, *when $d_{\mathrm{FW}}(t) = 0$ for $t \geq t_{\mathrm{FC}}$.*

**Proof:** From equation (3.16) and (3.21), we have the proved the property 1 in Laplace domain. Assuming $r_{\mathrm{MAX}}(t)$ is constant, the time-domain inequality of property 1 is given from equation (3.21), as

$$0 \leq e(t) \leq r_{\mathrm{MAX}}(T_K + T_{\mathrm{BW}}) \quad .$$

In equation (3.15), we have also shown that the inverse Laplace transform of equation (3.13) is given by

$$e(t) = \int_{t-T_{\mathrm{BW}}}^{t} d_{\mathrm{FW}}(\tau)d\tau + \int_{0}^{t-T_{\mathrm{BW}}} e^{-\frac{\tau}{T_{\mathrm{K}}}} d_{\mathrm{FW}}(t - T_{\mathrm{BW}} - \tau)d\tau.$$

The first integral of equation (3.15) is equal to 0 when $d_{\mathrm{FW}}(t) = 0$ for $t = t_{\mathrm{FC}} + T_{\mathrm{BW}}$. According to equation (3.13), second integral is due to a first order system with negative pole at $p = -\frac{1}{T_K}$, followed by delay equal to $t_{\mathrm{BW}}$. Thus, at $t \geq (t_{\mathrm{FC}} + T_{\mathrm{BW}})$, the second integral is asymptotically driven to zero with a time-constant of $T_K$, as shown before. More specifically, from equation (3.26), the error compensation transfer function has a negative pole at $p = -\frac{1}{T_K}$ and is asymptotically driven to 0, without any overshoot or oscillations, with a time constant of $T_K$. So, the property 2 is proved.

**Theorem 3.3.** *Given the initial condition $q(0) = 0, d_{\mathrm{FW}}(t) = 0$ for $[0 \leq t \leq (T_{\mathrm{FW}} + T_{\mathrm{BW}})]$ and $d_{\mathrm{BW}}(t) = 0$ for $[0 \leq t \leq T_{\mathrm{BW}}]$, if $T_K > 0$ and $r_{\mathrm{MAX}}(t)$ is constant, then the following properties hold true:*

1. $r_{\mathrm{MAX}}(t) \leq \frac{S}{T_{\mathrm{FW}}+T_{\mathrm{BW}}+T_K}$, *i.e. the maximum queue length is bound by the time-delays and maximum throughput.*

2. $0 \leq q(t) \leq S$, *i.e. instantaneous queue length is bound by maximum buffer length of $S$.*

3. *When congestion situation terminates, the $r_{\mathrm{DES}}(t)$ and $r_{\mathrm{REQ}}(t)$ are driven to $r_{\mathrm{MAX}}(t)$, which means the controller must retain the target rates whenever there is no congestive disturbance.*

**Proof:** From equation (3.18), it is evident that in time-domain the maximum buffer length is limited by

$$S = r_{\mathrm{MAX}}(t)\,[T_K + T_{\mathrm{BW}} + T_{\mathrm{FW}}] \quad .$$

Realistic implementation of the proposed control system relies on the assumption of the buffer length. As such, equation (3.18) sets the upper bound of the buffer length. However, for practical limitations, the physical interpretation of the buffer length comes from the fact that minimum $r_{\mathrm{MAX}}(t)$ is the rate at which the buffer can get full with the maximum delay of $(T_K + T_{\mathrm{BW}} + T_{\mathrm{FW}})$. This proves property 1 and property 2.

From equation (3.27), it is evident that when there is no forward disturbance, the error-controller asymptotically drives the $r_{\mathrm{DES_{ERR}}}$ to 0 with a time constant $T_{\mathrm{K}}$, given by

$$r_{\mathrm{DES_{ERR}}}(t) = \frac{1}{T_K}\int_0^t e^{-\frac{\tau}{T_K}} d_{\mathrm{FW}}(t-\tau)d\tau.$$

Since the compensation of $r_{\mathrm{DES}_{\mathrm{ERR}}}$ is only through $r_{\mathrm{DES}}(t)$ and not through $r_{\mathrm{REQ}}(t)$, the controller tracks the optimal data throughput according to IMC principles [18]. the closed-loop model strictly follows the target control to be achieved through meeting the control objective 3. This proves property 3.

### 3.3.2 Time-delay Performance

In order to analyze the robust stability of the interconnected closed-loop system derived in Sec. 3.2, time-delay analysis can be done. Since priori knowledge on the upper bounds would be known due to time-division multiplexed request assignment cycle, delay dependent time-domain analysis would suffice to illustrate the time-delay robustness. This kind of approach is based on various Lyapunov-Krasovski functionals and Lyapunov-Razumikhin functions [17]. However, in this research, owing to the fact that the time-delays are explicitly known by the design constraints in the master station, the system stability can be tuned by a proper choice of these delays during the design phase. In Chapter 5, several simulations are performed to assess the system stability under varied time-delays. It will be seen that the system performance is particularly affected by backward delay and not by forward delay due to equations (3.10) and (3.24). However, the effect of forward delay will have an impact in setting the maximum buffer size according to equation (3.14).

# Chapter 4

# Digital-filter Based Design

Based on the control system derived in Chapter 3, a digital filter-based model is derived in this chapter. With stable filter modeling, the reference controller can be implemented by the reference filter and the error controller can be implemented by the error filter, respectively. While doing necessary transformation, stability is strictly maintained, conforming to control objective 3. A complete schematic diagram is presented at the end of this chapter.

## 4.1 Discretization

In order to deduce a digital-filter based solution, the overall system derived in Sec. 3.1 is first discretized by sampling with a period of $T_s s$. With the high-speed large scale circuits available today, the sampling period can be made granular enough so that the forward and backward delays, $T_{\text{BW}}$ and $T_{\text{FW}}$, can be considered as integer multiples of $T_s$, $T_{\text{BW}} = \alpha T_s$ and $T_{\text{FW}} = \beta T_s$, where $\alpha$ and $\beta$ are non-negative integers. Also, since $T_K >> 2T_s$ and $T_K = \gamma T_s$, where $\gamma$ is also non-negative integer, from Nyquist theorem, it follows that the discretization will mimic the continuous time system, without any distortion or loss [23]. The fact that the forward and backward delays are explicitly known due to time-division multiplexed request assignment cycle, the choice of the sampling time can be decided beforehand. Following the discretized model, the continuous time system model shown in Sec. 3.2.1 and in Sec. 3.2.2 can be modeled by digital filters as shown in the following sections.

## 4.2 Reference Filter

Based on the reference controller in Sec. 3.2.1, a reference filter is found in this section using transformations, as required. Based on the desired and required reference rate, $r_{\mathrm{DES}_{\mathrm{REF}}}(s)$ and $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$, shown in equations (3.7) and (3.8), transformed and redefined rates in z-domain are given by

$$r_{\mathrm{DES}_{\mathrm{REF}}}(z) = r_{\mathrm{MAX}}(z) \quad , \tag{4.1}$$

$$r_{\mathrm{REQ}_{\mathrm{REF}}}(z) = r_{\mathrm{MAX}}(z)z^{-\alpha} - d_{\mathrm{BW}}(z) \quad . \tag{4.2}$$

Since $q_{\mathrm{REF}}(z)$ is necessarily the integrated difference between the desired input and required output rates, it can be expressed in terms of $r_{\mathrm{MAX}}(z)$ and $d_{\mathrm{BW}}(z)$ as shown in equation (3.11) with necessary transformations, i.e.

$$q_{\mathrm{REF}}(z) = \frac{1 - z^{-(\alpha+\beta)}}{z-1} r_{\mathrm{MAX}}(z)T_s + \frac{z^{-\beta}}{z-1} d_{\mathrm{BW}}(z)T_s \quad . \tag{4.3}$$

The expression for the reference queue length in equation (4.3) can be simplified as

$$q_{\mathrm{REF}}(z) = \underbrace{\frac{z^{-1} - z^{-(\alpha+\beta+1)}}{1 - z^{-1}}}_{\text{IIR Filter 1}} r_{\mathrm{MAX}}(z)T_s + \underbrace{\frac{z^{-(\beta+1)}}{1 - z^{-1}}}_{\text{IIR Filter 2}} d_{\mathrm{BW}}(z)T_s \quad . \tag{4.4}$$

Equation (4.4) models the reference filter by two infinite impulse response (IIR) filter transfer functions that contribute to setting the $q_{\mathrm{REF}}(z)$ from $r_{\mathrm{MAX}}(z)$ and $d_{\mathrm{BW}}(z)$ over one time sample, $T_s$. Also, according to equation (4.4), the first IIR filter must have a minimum of $(\alpha + \beta + 1)$ taps, while the second IIR filter must have more than $(\beta + 1)$ taps in order to be allow time-delay effects. The fact that they add together also requires that $r_{\mathrm{MAX}}(z)T_s$ and $d_{\mathrm{BW}}(z)T_s$ have same tap length, i.e. a minimum length of $(\alpha+\beta+1)$ taps. The reference model is a direct consequence of the reference controller and is stable. This is due to pole on the unit circle in the z-plane as shown in equation (4.3). The term $T_s$ appear in the expression along with $r_{\mathrm{MAX}}(z)$ and $d_{\mathrm{BW}}(z)$ due to inherent conversion of data rate in (in bps) to buffer length in (in bits). This can easily implemented using a single tap multiplier with the data line.

## 4.3 Error Filter

As seen in Sec. 3.2.2, based on the error between instantaneous queue length and $q_{\mathrm{REF}}(z)$, and found in Sec. 4.2, the error filter must generate the compensation for $d_{\mathrm{FW}}(z)$ as $r_{\mathrm{DES}_{\mathrm{ERR}}}(z)$ and $r_{\mathrm{REQ}_{\mathrm{ERR}}}(z)$. From equations (3.24) and (3.25), $r_{\mathrm{DES}_{\mathrm{ERR}}}(z)$ and $r_{\mathrm{REQ}_{\mathrm{ERR}}}(z)$ can be expressed as

$$\frac{r_{\mathrm{DES}_{\mathrm{ERR}}}(z)}{e(z)} = \frac{\frac{1}{\gamma Ts}}{1+\frac{1}{\gamma(z-1)}(1-z^{-\alpha})} = \underbrace{\frac{\frac{1}{T_s}(1-z^{-1})}{\gamma+(1-\gamma)z^{-1}-z^{-(\alpha+1)}}}_{\text{IIR Filter 3}} \quad (4.5)$$

$$\frac{r_{\mathrm{REQ}_{\mathrm{ERR}}}(z)}{e(z)} = 0. \quad (4.6)$$

where $e(z)$ is the z-domain queue error and can be expressed from equations (3.12) as $e(z) = q(z) - q_{\mathrm{REF}}(z)$. The transfer function shown in equation (4.5) is an IIR filter with $(\alpha+1)$ feedback taps and is stable. The term $\frac{1}{T_s}$ appears in the numerator due to conversion between a rate (in bps) and queue length (in bits). As before, no compensation is needed to reduce the required output reference rate and as such the transfer function between $r_{\mathrm{REQ}_{\mathrm{ERR}}}(z)$ and $e(z)$ is zero. Hence, actual desired and required rates, $r_{\mathrm{DES}}(z)$ and $r_{\mathrm{REQ}}(z)$ can be expressed as

$$r_{\mathrm{DES}}(z) = r_{\mathrm{DES}_{\mathrm{REF}}}(z) - r_{\mathrm{DES}_{\mathrm{ERR}}}(z), \quad (4.7)$$

$$r_{\mathrm{REQ}}(z) = r_{\mathrm{REQ}_{\mathrm{REF}}}(z) - r_{\mathrm{REQ}_{\mathrm{ERR}}}(z). \quad (4.8)$$

These rates directly affect the input and output rates that the system might have. However, the input and output rates cannot be known beforehand because disturbance is generated on some criterion externally. The actual input and output rates are found out from the transformed time-delay models as

$$r_{\mathrm{IN}}(z) = r_{\mathrm{DES}}(z)z^{-\alpha} - d_{\mathrm{BW}}(z), \quad (4.9)$$

$$r_{\mathrm{OUT}}(z) = r_{\mathrm{REQ}}(z)z^{-\beta} - d_{\mathrm{FW}}(z). \quad (4.10)$$

The delays in equations (4.9) and (4.10) can be modeled by finite impulse response (FIR) filters of $(\alpha + 1)$ and $(\beta + 1)$ taps respectively, with all but the final tap set to one. Also, to make sure the rates do not fall below zero due to external disturbance rates, limiters can be used. From equation (3.2) and input and output rates shown above, the instantaneous queue length, $q(z)$, will thus grow as

$$q(z) = \frac{T_s}{z-1}\left[r_{\mathrm{IN}}(z) - r_{\mathrm{OUT}}(z)\right] = T_s \underbrace{\frac{z^{-1}}{(1-z^{-1})}}_{\text{IIR filter 4}} \left[r_{\mathrm{IN}}(z) - r_{\mathrm{OUT}}(z)\right]. \qquad (4.11)$$

The IIR filter implementation (functioning as an integrator) in equation (4.11) needs only two taps and is marginally stable, since only one pole exists on the unit circle. However, it's stability can be improved during the feedback process. The difference between reference and instantaneous queue length would thus be fed to the filter-based control system after every $T_s$ seconds, but the actual compensation for forward congestion must wait for at least $\beta$ samples. An integrated filter-based solution to the proposed congestion controller is shown in Figure 4.1.

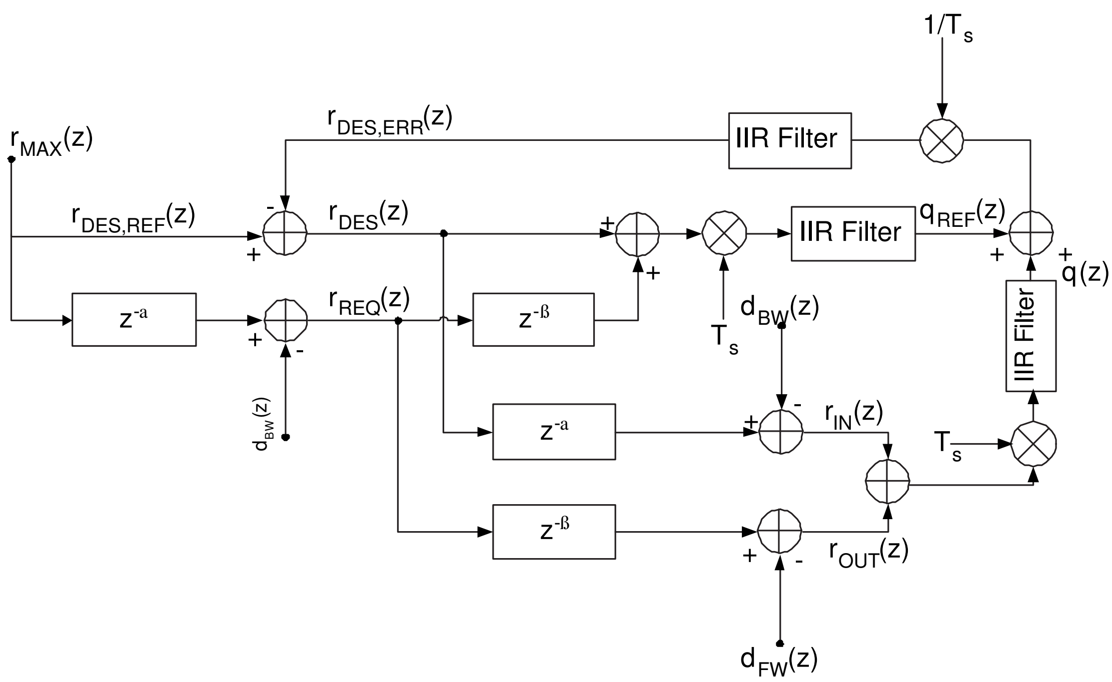


FIGURE 4.1: Integrated filter-based congestion controller

## 4.4 Filter Responses and Comments

The filter based model shown in Figure 4.1 is the complete integrated solution to the control of congestion. The external inputs are: maximum throughput rate, $r_{\text{MAX}}(z)$, backward disturbance, $d_{\text{BW}}(z)$ and forward disturbance, $d_{\text{FW}}(z)$. The filter-based model handles the disturbances separately but in an integrated manner. The former affects the reference filter, which functions as a open loop IMC controller for modeling reference queue length, while the later affects the actual input and output rates through a closed-loop filter based controller to mitigate the effect of the disturbances.

While deriving, it has been extensively shown that the filters are stable considering the pole and zero locations and are easy and straightforward for implementation. But it is also important that the values of $\alpha$, $\beta$ and $\gamma$ are carefully chosen, otherwise numerator terms in IIR filters 1 and 2 with rapid change of values may be inflicted as a faster compensation requirement at the closed-loop error controller. Faster compensation is particularly hard to achieve since the situation worsens due to lag in compensation on every forward and backward time delay interval. In Figure 4.2 and Figure 4.3, the effect of change of times delays on the two frequency response of the filters are shown similar to that in digital signal processing (DSP) circuits. While in this section, the frequency domain effects are investigated, the effect of change of time-delays on the time domain performance are illustrated more comprehensively in Chapter 5.

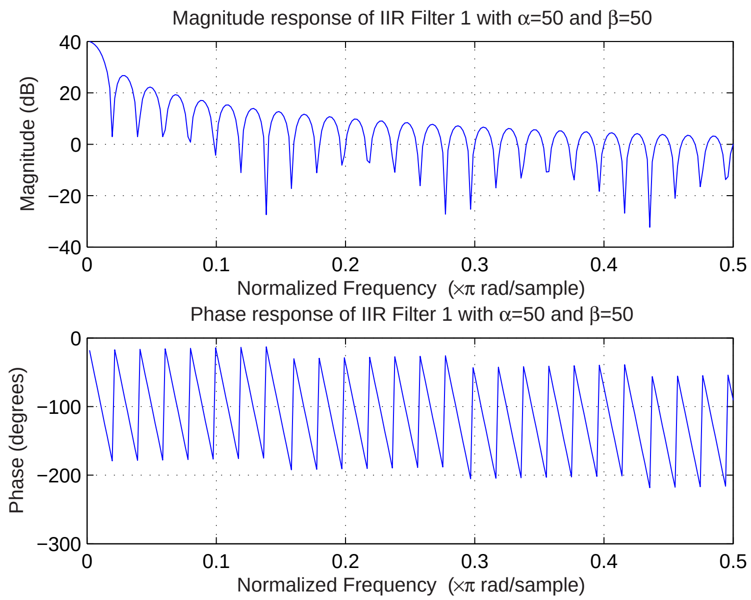


FIGURE 4.2: Effect of $\alpha = 50$ and $\beta = 50$ on the IIR Filter 1 for $r_{\text{MAX}}(z)T_s$

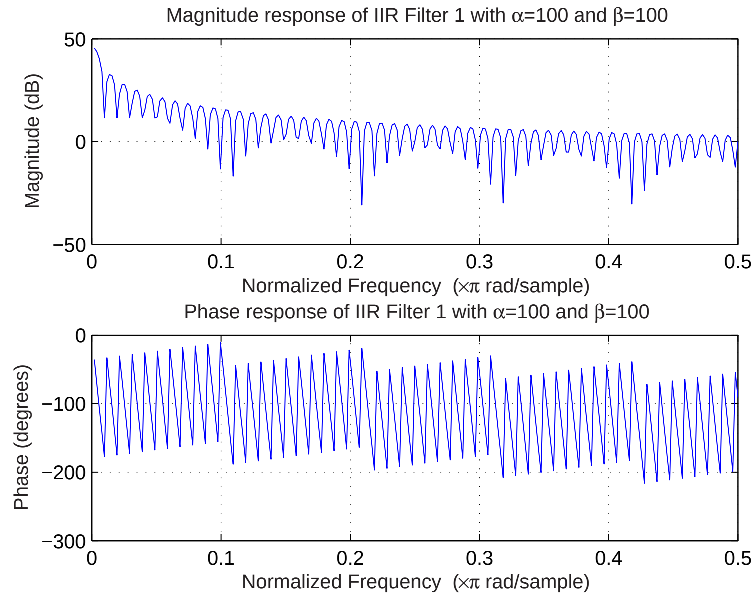


FIGURE 4.3: Effect of $\alpha = 100$ and $\beta = 100$ on the IIR Filter 1 for $r_{\text{MAX}}(z)T_s$

In our assumptions, since $r_{\text{MAX}}(z)T_s$ is constant and does not change, the magnitude response does not reveal any useful information from Figure 4.2 and Figure 4.3. However with the increase of $\alpha$ and $\beta$, the phase changes faster. The IIR filter 1 necessarily accumulates sample-by-sample differences between maximum throughput rate with delayed version of the same by $(\alpha + \beta)$ samples. Similar to IIR filter 1, Figure 4.4 and Figure 4.5 demonstrate the contribution of backward delay and effect of such delay in the reference filter.

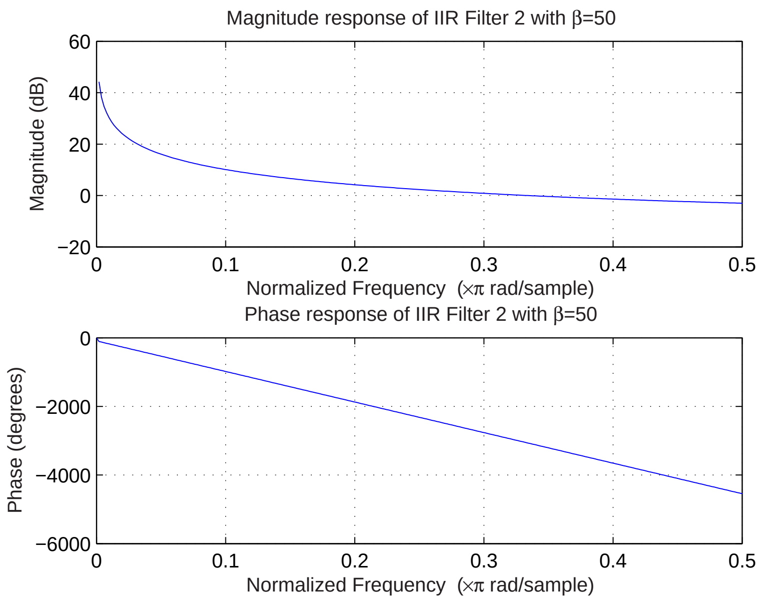


FIGURE 4.4: Effect of $\beta = 50$ on the IIR Filter 2 for $d_{\text{BW}}(z)T_s$

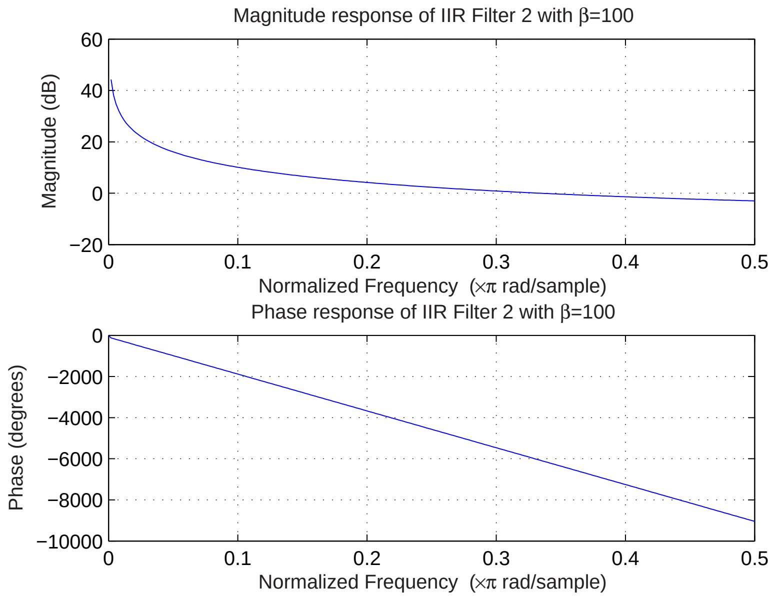


FIGURE 4.5: Effect of $\beta = 100$ on the IIR Filter 2 for $d_{\mathrm{BW}}(z)T_s$

As shown in Figure 4.4 and Figure 4.5, the magnitude of the backward disturbance is not the same over all frequencies rather the filter acts more like a low pass amplifier: the slower changes of the backward disturbance rate over a sample time is amplified and higher changes are attenuated. As far as the delay is concerned, it is asymptotically increasing in values and the slope depends on how high or low the delays are. With larger delay, the increase of the phase is higher and steeper, while for smaller delays this is rather smaller, as well. These are DSP analogies that show that the proposed filter based scheme works exactly in the same way as the continuous time model shown in Chapter 3.

In Figure 4.6, 4.7 and 4.8, the effects of change of values of $\alpha$ and $\gamma$ are shown. Since $\gamma$ is the constant related to setting of the maximum buffer length, the only effect of this would be in changing the upper limit, while change of $\alpha$ changes the how quickly the compensation can be done. The change of $\alpha$ from 50 to 100 in Figure 4.6 and Figure 4.7 demonstrate the fact that with higher $\alpha$ the filter is able to amplify the rate of change of backward disturbance even faster. However as far as the phase response is concerned, the phase experiences higher damping for faster rate of change with higher initial phase. This is due to the integration like operation carried out in the time-domain. As demonstrated in Figure 4.8, with the increase of $\gamma$, the amplification at higher rate of change of backward disturbance is increased and phase damping is reduced as the rate experiences sharp change.

In all the simulations carried out in Chapter 5, only the time domain effects and

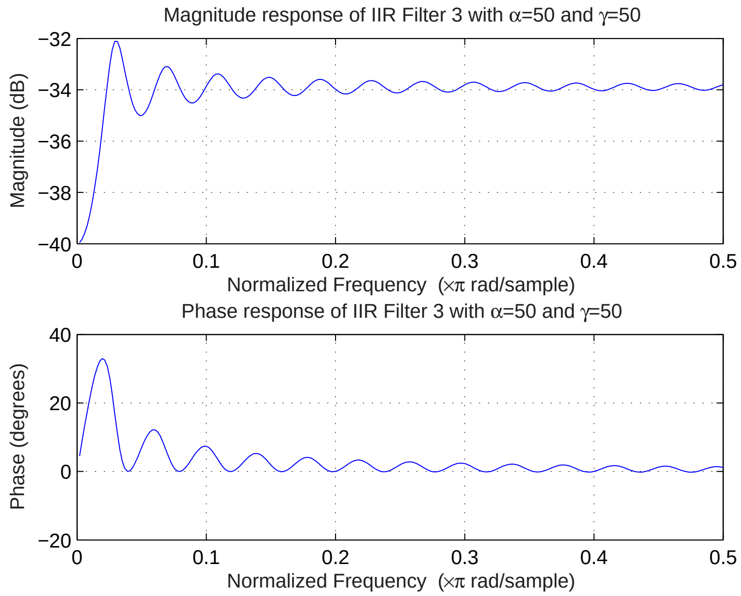


FIGURE 4.6: Effect of $\alpha = 50$ and $\gamma = 50$ on the IIR Filter 3 for error compensation transfer function

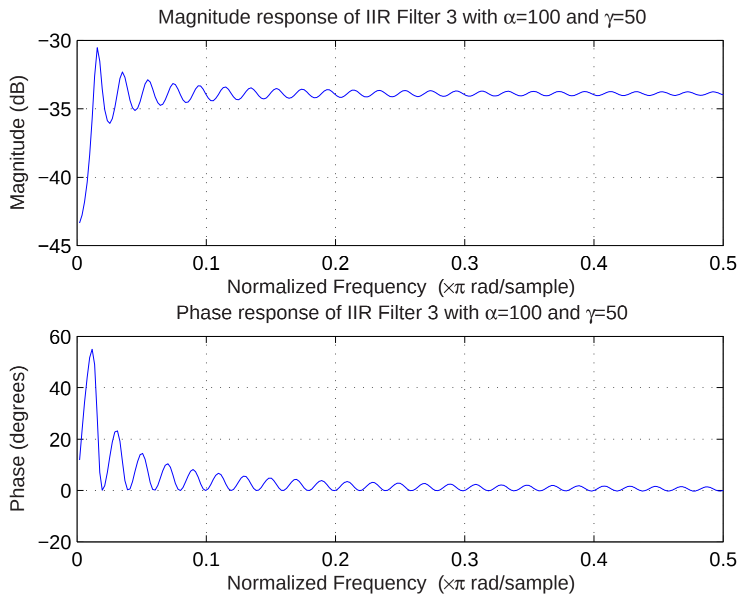


FIGURE 4.7: Effect of $\alpha = 100$ and $\gamma = 50$ on the IIR Filter 3 for error compensation transfer function

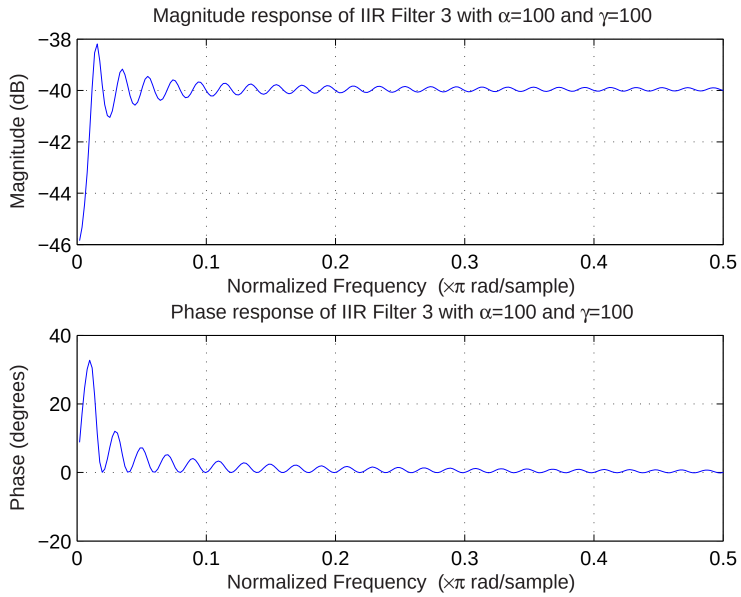


FIGURE 4.8: Effect of $\alpha = 100$ and $\gamma = 100$ on the IIR Filter 3 for error compensation transfer function

not the frequency domain effects are demonstrated that may affect the system. But it is important to have rate of change effects taken into consideration in order to modulate the system stability, because in real systems, these changes come as rectangular short-duration signals. In all the filters demonstrated above, the system does not have any instability problems for any particular rate of change that the system might experience. From control-theoretic point of view, this is particularly important since stability is set as the third but most important objectives in Chapter 3.

# Chapter 5

# Simulations and Results

Assuming discrete fluid flow approximation for data in the WLAN for the controller, appropriately chosen simulation based tests are undertaken to assess the achievable performance of the system described in Sec. 4.1. In the simulations, the nodes in the WLAN are now classified according to their functions in the network. Source and destination nodes are the participating transmitting and receiving nodes, while bridges are the connecting nodes between the clusters. The model derived in Sec. 3.1 and in Sec. 4.1, can be directly placed in the bridge controller. However, since the source controller does not make any $r_{\mathrm{DES}}(z)$ request, the time-delay model for the backward delay is thus not needed. Also for the destination controller, the time-delay model for the forward delay is not needed since it does not make any outbound request in the form of $r_{\mathrm{REQ}}(z)$. In this chapter, the proposed filter based model in Chapter 4 is simulated in different scenarios to illustrate the effects of time-delays and effect of disturbances by MATLAB Simulink. Also with an integrated WLAN scenario modeled in OPNET, the discrete event simulator by MIL3, simulations have been carried out to investigate the performance of the integrated system.

## 5.1 Effects of Congestive disturbances

Based on the derivations in Sec. 4.1, the following parameter values are assumed for simulation to illustrate the effects of disturbances: $T_s = 0.001s$, $r_{\mathrm{MAX}}(t) = 11Mbps$, $T_{\mathrm{BW}} = 100T_s$, $T_{\mathrm{FW}} = 200T_s$, $T_{\mathrm{K}} = 100T_s$. The choice of $T_s$ is such that forward and backward delays can be effectively implemented using delay filters, as mentioned in Chapter 4. As such, $r_{\mathrm{MAX}}(z)$, $d_{\mathrm{BW}}(z)$, $d_{\mathrm{FW}}(z)$, $r_{\mathrm{IN}}(z)$ and $r_{\mathrm{OUT}}(z)$ are provided with buffer lengths of 600 each for the delays which can be implemented by filters shown in Sec. 4.1. In all the following simulations, the initial queue length is assumed to be of zero length and the units in the following descriptions are avoided due to discretization but implicitly understood.

### 5.1.1 Effects of Backward Congestion

Backward disturbance is generated when the master station acknowledges with input rate, $r_{\mathrm{IN}}(t)$, different from the original desired input rate, $r_{\mathrm{DES}}(t)$, requested by the nodes. Thus, in order to assess the effect of backward disturbance, it is assumed that the master station cannot allow any input to be taken against the desired input requests during $t = 0.5s$ to $t = 2.5s$ and it allows whatever output requests are made during this time. These mean that the system will be experiencing a backward disturbance of $d_{\mathrm{BW}}(z) = 11 \times 10^6$ and $d_{\mathrm{FW}}(z) = 0$ during the interval considered. The congestion disturbances are shown in Figure 5.1 along with the respective desired, required, input and output rates. Due to equation (4.1), the reference desired rate will always be set to $r_{\mathrm{DES}}(z) = 11 \times 10^6$, while due to equation (4.2), $r_{\mathrm{REQ}}(z) = 11 \times 10^6$ after a time-delay of $T_{\mathrm{BW}} = 0.1s$. From $t = 0.1s$ to $t = 0.5s$, the desired and required rates would be sustained until a backward disturbance happens at $t = 0.5s$. Also during this interval $r_{\mathrm{IN}}(z) = 11 \times 10^6$ after $r_{\mathrm{DES}}(z)$ is allowed by the master station at $t = 0 + T_{\mathrm{BW}}s = 0.1s$. Since $r_{\mathrm{REQ}}(z)$ is already delayed by $0.1s$, due to the forward delay loop, $r_{\mathrm{OUT}}(z) = 11 \times 10^6$ after $t = 0.1 + T_{\mathrm{FW}} = 0.3s$. After $t = 0.5s$, since no input is being allowed, $r_{\mathrm{IN}}(z) = 0$ against $r_{\mathrm{DES}}(z) = 11 \times 10^6$ and as a result, $r_{\mathrm{OUT}}(z) = 0$ after a forward delay of $200T_s = 0.2s$. As soon as backward congestion is cleared at $t = 2.5s$, $r_{\mathrm{DES}}(z) = 11 \times 10^6$ and as such, $r_{\mathrm{IN}}(z)$ is revived back to $r_{\mathrm{MAX}}(z) = 11Mbps$ according to equation 3.4. This is necessarily control objective 2 stated in Chapter 3. The output rate, $r_{\mathrm{OUT}}(z)$, is also set to $r_{\mathrm{MAX}}(z) = 11 \times 10^6$, after a forward request to the master station is done without any disturbance being generated during $T_{\mathrm{FW}} = 200T_s s$. Typical simulation results which support these statements are given in Figure 5.1.

As far as the queue lengths are concerned, the growth of the reference queue length is directly affected by the backward disturbance according to equation (4.3). At $t = 0.1s$ onwards, during this interval of $T_{\mathrm{FW}} = 200T_s = 0.2s$, $q_{\mathrm{REF}}(z)$ grow linearly according to integral function of the filters shown in equation 4.3. Since there is no forward congestion, the error compensation is immediate to make sure $q(z)$ follows $q_{\mathrm{REF}}(z)$ as shown in Figure 5.1. During the $0.2s$ interval the reference

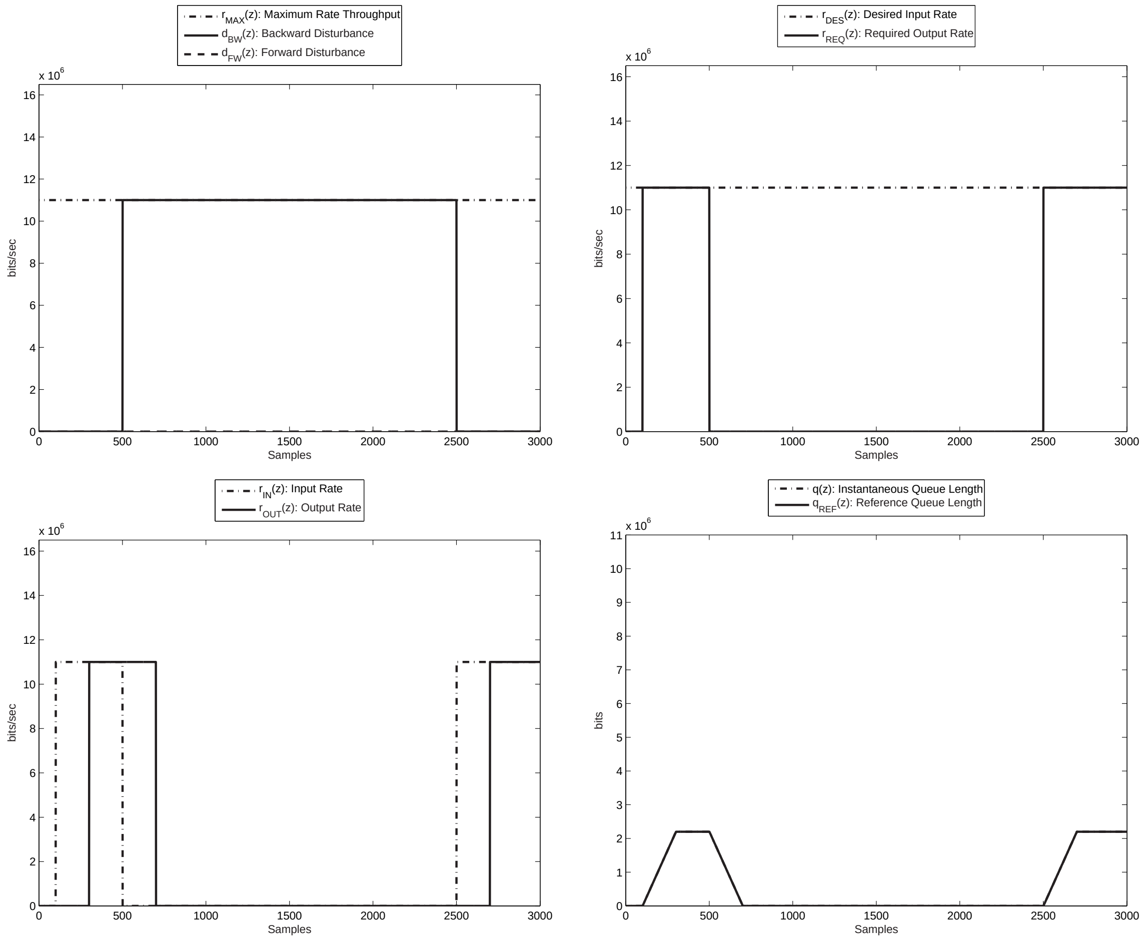


FIGURE 5.1: Simulation 1: Effect of backward disturbance, $T_{\mathrm{FW}} = 0.2s$, $T_{\mathrm{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

and instantaneous queue lengths grow up to $11 \times 10^6 \times 0.2 = 2.2 \times 10^6 bits$. At $t = 0.3s$ and onwards, this is maintained until the buffer has some opportunity to dispatch some of its stored data due to backward disturbance at this time. However due to forward-loop delay and $r_{\mathrm{IN}}(z) = 0$ and $r_{\mathrm{IN}}(z) = 11 \times 10^6$ at $t = 0.5s$ onwards, the queue is emptied at a rate of $11 \times 10^6$ within $0.2s$. Since no transmissions take place from $t = 0.7s$, the queue remains empty until $t = 2.5s$, when backward disturbance is cleared. At this time again, the queue starts to build up to $2.2 \times 10^6 bits$ due to forwards-loop delay between $r_{\mathrm{IN}}(z)$ and $r_{\mathrm{OUT}}(z)$. It is to be noted that due to no closed-loop error feedback, the reference and error filter employs perfect IMC principle and the reference and instantaneous queue lengths are exactly the same.

### 5.1.2 Effects of Forward Congestion

Similarly to backward disturbance, forward congestion is generated when the master station acknowledges with output rate, $r_{\mathrm{OUT}}(t)$, different from the original required output rate, $r_{\mathrm{REQ}}(t)$, requested by the nodes for an outbound transmission request. In order to assess the effect of forward disturbance, it is assumed that the master station cannot allow any output to be taken against the required output rate requests from $t = 0.5s$ to $t = 2.5s$, while it allows whatever input requests are made during this time. In this case, the system will experience a forward disturbance of $d_{\mathrm{FW}}(z) = 11 \times 10^6$ and $d_{\mathrm{BW}}(z) = 0$ during this time. The congestion disturbances are shown in Figure 5.2 along with the respective desired, required, input and output rates. The desired input rate, $r_{\mathrm{DES}}(z)$ will be set the maximum $11 \times 10^6$ until the forward disturbance appears as shown previously in equation (4.1). Since no compensation is done on the required output rate, it will also be set to $11 \times 10^6$ after a backward loop delay of $0.1s$ that would be taken in the backward request process. At $t = 0.5s$, as the forward congestion takes place, due to equation (4.5), the desired rate will be reduced due to error compensation. This compensation is exponential and increases with a time constant of $T_{\mathrm{K}}$. At approximately $t = 1s$, the compensation would cause the desired rate fall at zero. Since there is no backward congestion assumed, the input rate would follow exactly, except for a time-lag of backward-loop delay. Also as a consequence of forward congestion, output rate will be $11 \times 10^6 bps$ from $t = 0.3s$ to $t = 0.5s$ owing to forward loop delay. The desired rate, require rate, input and outputs rates are shown in Figure 5.2. Typical simulation results which support these statements are given in Figure 5.2. As soon as forward congestion is cleared at $t = 2.5s$, $r_{\mathrm{DES}}(z)$ regains back to $11 \times 10^6 bps$ due to exponential reduction in compensation. Since there is no compensation, $r_{\mathrm{IN}}(z)$ is is still maintained at $r_{\mathrm{MAX}}(z) = 11Mbps$. The desired rate revival is almost immediate following a time constant of $T_{\mathrm{K}}$ as this is necessarily control objective 2. The output rate, $r_{\mathrm{OUT}}(z)$, is also set to $r_{\mathrm{MAX}}(z) = 11 \times 10^6$ immediately with the requested output rate as no forward disturbance exists after $t = 2.5s$. This is also illustrated in Figure 5.1.

As far as the queue lengths are concerned, the growth of the reference queue length is directly affected by the forward disturbance according to equation (4.3). At $t = 0.1s$ onwards, during an interval of $T_{\mathrm{FW}} = 200T_s = 0.2s$, $q_{\mathrm{REF}}(z)$ grow linearly according to integral function of the filters shown in equation 4.3. The reference and instantaneous queue lengths grow up to $11 \times 10^6 \times 0.2 = 2.2 \times 10^6 bits$. At $t = 0.3s$ and onwards, this is maintained up to $t = 0.5s$ until the forward

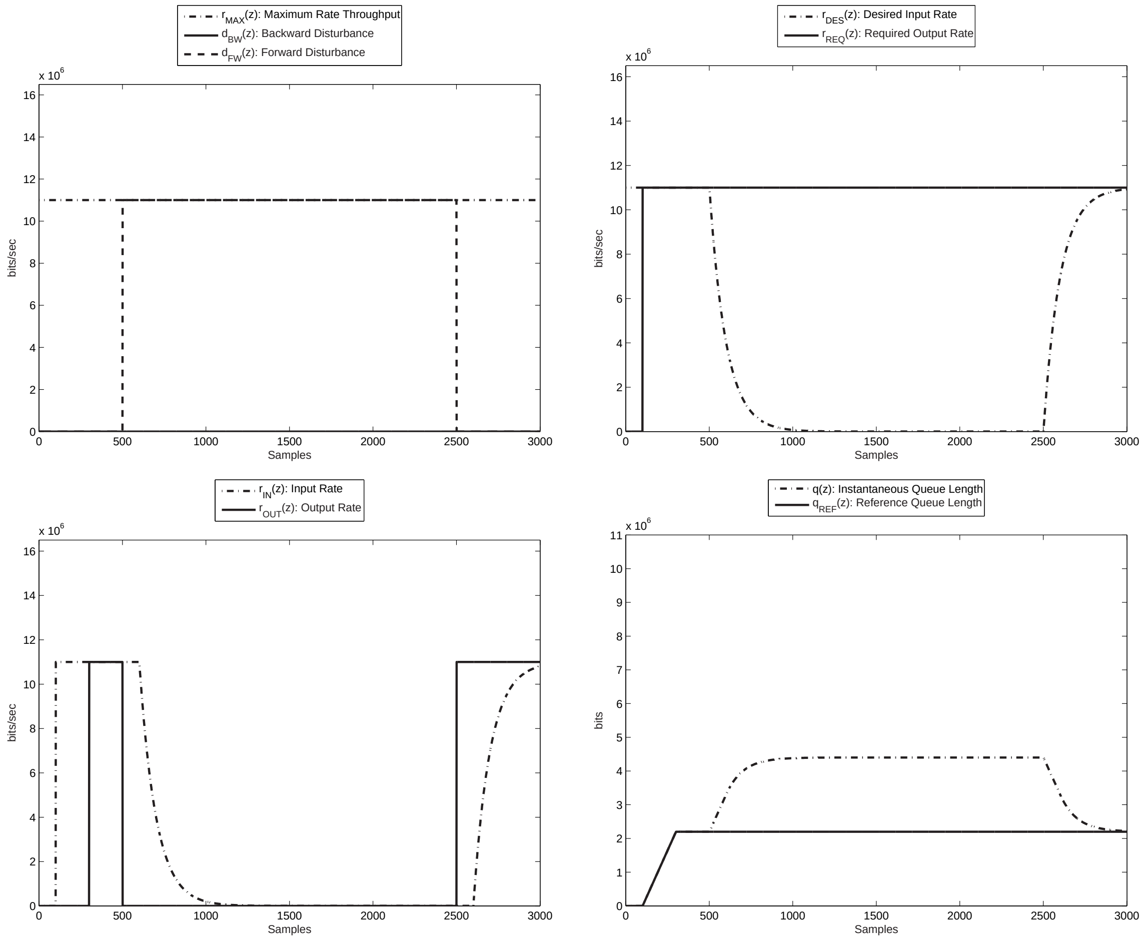


FIGURE 5.2: Simulation 2: Effect of forward disturbance, $T_{\mathrm{FW}} = 0.2s$, $T_{\mathrm{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

disturbance happens. At $t = 0.5s$ onwards, due to $0.1s$ lag the instantaneous queue starts to build up exponentially with a time constant of $T_{\mathrm{K}}$ and after approximately $0.8s$, the queue length is sustained until forward congestion is over. At $t = 2.5s$, the forward congestion is cleared and the input rate gradually starts to build up with time delay of $T_{\mathrm{BW}} = 0.1s$ again. This causes the queue to come back to the reference queue length value with the same time constant. During the forward disturbance, the control actions are critical dual IMC problem. In the closed-loop error controller a rather loosely-coupled IMC controller is used to ensure the system is asymptotically stable, while the queue limits are maintained by control objective 1.

### 5.1.3 Effects of Combined Congestion

In a real ad hoc networks, normally the disturbances appear simultaneously with a with a time difference of $T_{\text{BW}}$ between the backward congestion and the forward congestion. This is due to the fact that the forward disturbance take place only when an input, which is also the reference for output request is acknowledged. Thus, in order to assess the effect of forward and backward disturbances, it is assumed that the master station can only allow input to be taken $5.5 \times 10^6 bps$ less than that of desired input requests during $t = 0.5s$ to $t = 2.5s$ and it allows output rate $5.5 \times 10^6 bps$ less than that of required output rate requests during $t = 0.6s$ to $t = 2.6s$ . These mean that the system will be experiencing a backward disturbance of $d_{\text{BW}}(z) = 5.5 \times 10^6$ and $d_{\text{BW}}(z) = 5.5 \times 10^6$ during the interval mentioned. The congestion disturbances are shown in Figure 5.3 along with the respective desired, required, input and output rates. As before, due to equation (4.1), the reference desired rate will always be set to $r_{\text{DES}_{\text{REF}}}(z) = 11 \times 10^6$, while due to compensation shown in equation (4.5), $r_{\text{DES}}(z)$ will fall exponentially down to $5.5 \times 10^6$ after a time-delay of $T_{\text{BW}} = 0.1s$. During the interval between $t = T_{\text{BW}} = 0.1s$ and $t = 0.5s$, the required rate will be set to the maximum $11 \times 10^6 bps$. The delay is due to the backward loop request that has to happen before the required rate is known to the system. Due to presence of backward disturbance in the system starting from $t = 0.5s$ to $t = 2.5s$, the required rate will fall down to $5.5 \times 10^6 bps$. Again, due to presence of forward disturbance in the system starting from $t = 0.6s$ to $t = 2.6s$, the desired rate will fall down to $5.5 \times 10^6 bps$ exponentially with the error filter feeding back the compensation as $r_{\text{DES}_{\text{ERR}}}(z)$. The corresponding desired and required rates are shown in the Figure 5.3. The impact of these desired and required rates are direct on the input and output rates. The input rate is related to the desired rate with a backward delay and the backward disturbance. The input rate fall down to $0bps$ in approximately the same time when the desired rate falls down to $5.5 \times 10^6 bps$. Since at $t = 2.5s$, the backward congestion is cleared, the desired and hence the input rate exponentially rise due to reduction in the exponential compensation. Similarly, output rate is directly affected by the required rate, with a time forward delay and forward disturbance coming into play. Since there is no compensation done on the required rate, this relations are straight forward and as shown in Figure 5.3.

As far as the queue lengths are concerned, the growth of the reference queue length is directly affected by the combined disturbance according to equation (4.3). At $t = 0.1s$ onwards, during this interval of $T_{\text{FW}} = 200T_s = 0.2s$, $q_{\text{REF}}(z)$ grow

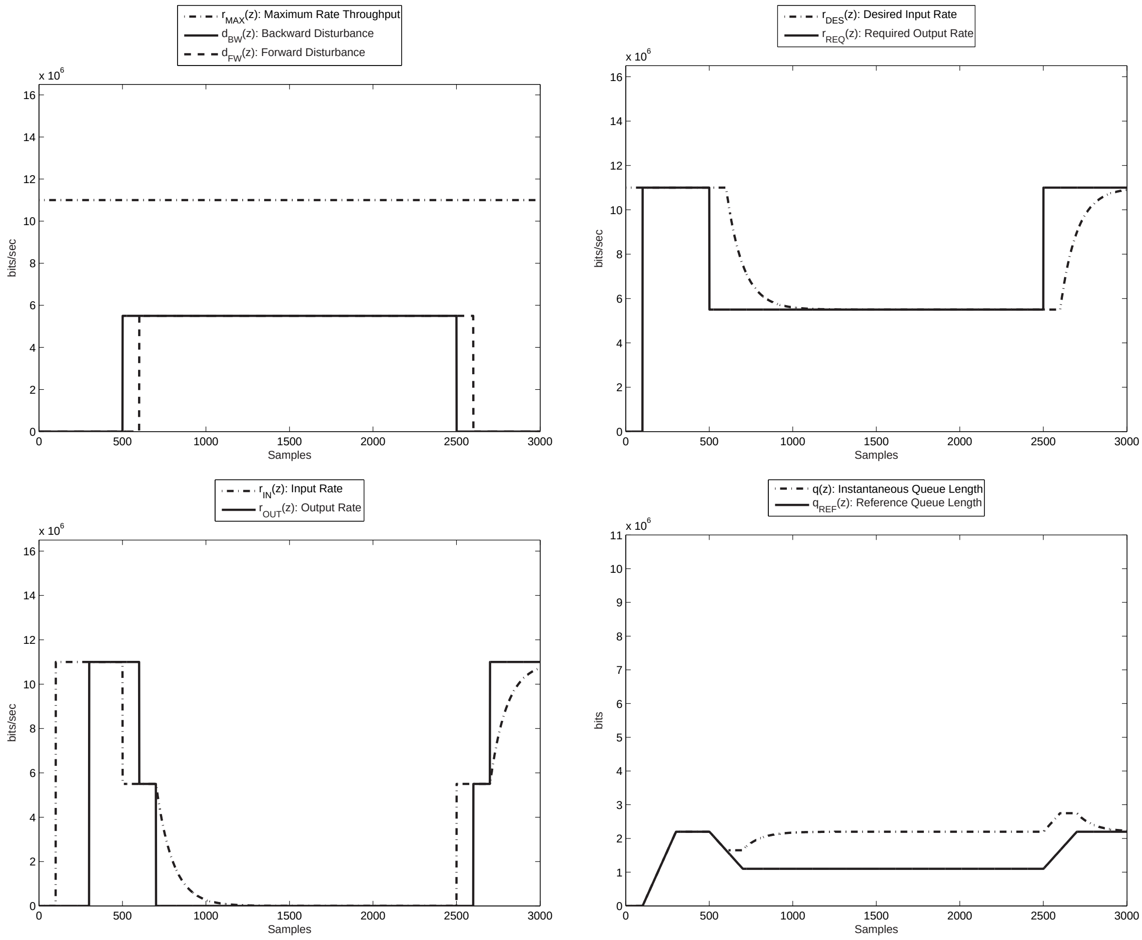


FIGURE 5.3: Simulation 3: Effect of combined disturbance, $T_{\mathrm{FW}} = 0.2s$, $T_{\mathrm{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

linearly according to integral function of the filters shown in equation 4.3. Since there is no forward congestion, the error compensation is immediate to make sure $q(z)$ follows $q_{\mathrm{REF}}(z)$ as shown in Figure 5.3. During the $0.2s$ interval the reference and instantaneous queue lengths grow up to $11 \times 10^6 \times 0.2 = 2.2 \times 10^6 bits$. At $t = 0.3s$ and onwards, both the reference and instantaneous queue are maintained at a constant level since there is no disturbance present. However due to the fact forward disturbance effect makes the input rate to be higher than the output rate, from $t = 0.6s$, the instantaneous slightly deviates from the reference queue length with an exponential rise to $11 \times 10^6 \times 0.2 = 2.2 \times 10^6 bits$. This continues until $t = 2.5s$, after which due to difference between the rates, both the queue lengths increase integrally for a duration of $0.1s$. From $t = 2.6s$, since both the input and output rates reach the maximum rate $11 \times 10^6$, both the queues level up and follow the open-loop IMC controller set values asymptotically as long as the forward disturbance does not happen.

## 5.2 Effects of Time-delays

Knowledge of bounded values of time delays are crucial to stability of time-delay systems [17]. In this system a time-division multiplexed demand assignment cycle at the master node is implemented and hence these delays are explicitly known. This is essentially the requirement for the control mission set in objective 3 in Chapter 3. However, for a strictly stable performance it is important to find out the best and minimum possible time-delay that can be fit into the design. In this section, the effect of change of forward and backward time-delays on the different rates and the queue lengths and on the system stability has been investigated.

To investigate the effects of the time delays only, a simple case is considered when the controller will be subjected to forward disturbance only since backward disturbance cannot affect the compensation and as such time delays effects are not reflected in the results. The forward and backward delays are set as $T_{\mathrm{FW}} = 0.2s$ and $T_{\mathrm{BW}} = 0.1s$. As shown in Figure 5.4, a simulation is performed as a reference for the later two cases where the forward and backward delays are increased to simulate how they affect the system performance. The system is considered as subjected to a forward disturbance of $11 \times 10^6 bps$ from $t = 1.0s$ to $t = 2.0s$. As expected and shown in Sec. 5.1, the desired rate exponentially drops to $0bps$ at $t = 1.0s$ with a time constant of $T_{\mathrm{K}}$. Also, the required rate would be set to maximum rate since no backward congestion takes place as shown in equation (4.2). Due to loop delays, the corresponding input and output rates are affected as shown in the Figure 5.4. Due to Difference between the input and output rate at and onwards $t = 1.0s$, the instantaneous queue length exponentially rises to $11 \times 10^6 \times (T_{\mathrm{FW}} + T_{\mathrm{K}} + T_{\mathrm{BW}}) = 4.4 \times 10^6 bits$. This is maintained until the congestion is cleared at $t = 2.0s$, when the rates rise exponentially and the instantaneous queue length follows the reference queue length.

### 5.2.1 Effects of Forward Delay

In the following simulation, the forward delay, $T_{\mathrm{FW}}$, has been increased from $0.2s$ to $0.5s$ to investigate the effect of increase of forward delay towards the system performance. As shown in Figure 5.5, increasing the forward delay simply increases

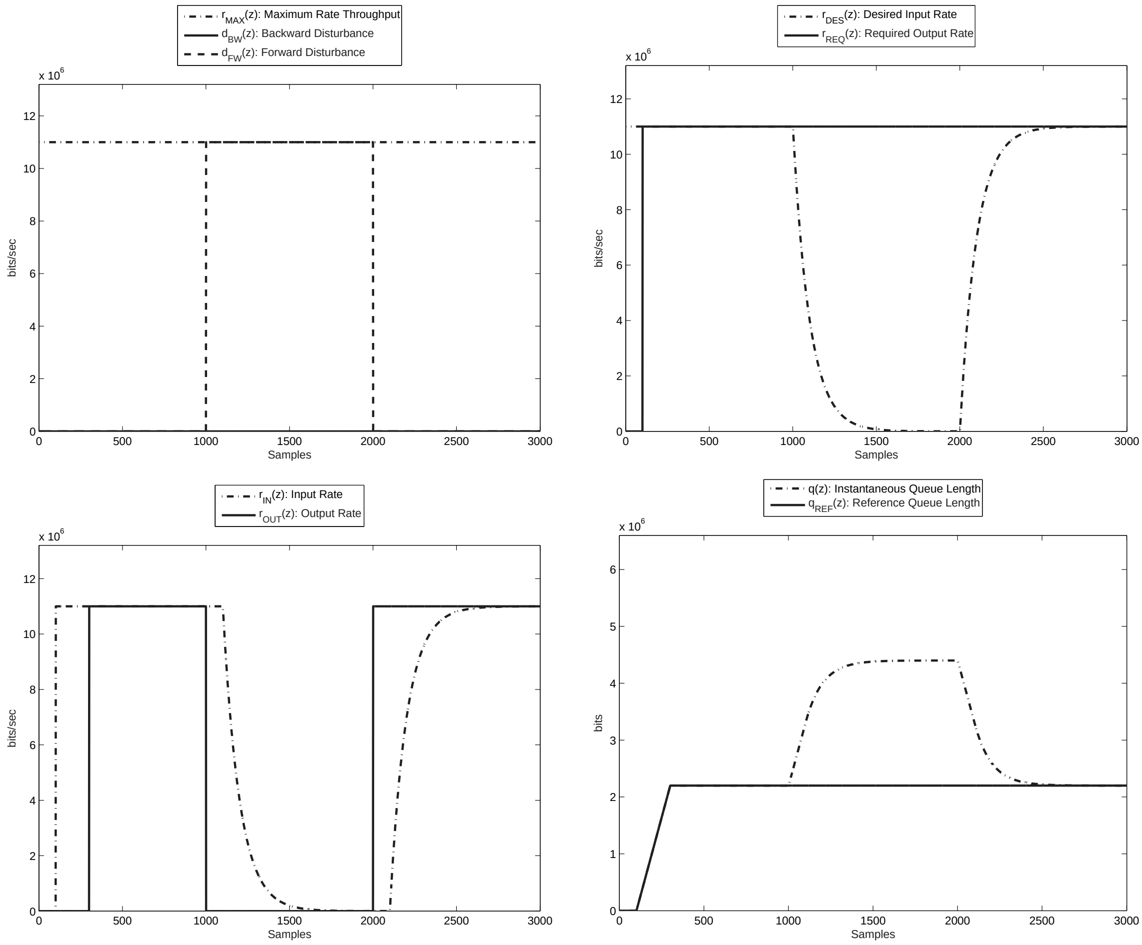


FIGURE 5.4: Simulation 4: Effect of delay, $T_{\mathrm{FW}} = 0.2s$, $T_{\mathrm{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

the reference queue length, $q_{\mathrm{REF}}(z)$, and hence the instantaneous queue length, $q(z)$. This causes the maximum instantaneous queue length to be $11 \times 10^6 \times (T_{\mathrm{FW}} + T_{\mathrm{K}} + T_{\mathrm{BW}}) = 7.7 \times 10^6 bits$. This is maintained until the congestion is cleared at $t = 2.0s$, when the rates rise exponentially and the instantaneous queue length follows the reference queue length, as shown in Figure 5.5.

### 5.2.2 Effect of Backward Delay

Backward delay is the most crucial element in system stability since it appears as a Smith predictor variable in the closed loop error controller as shown in equation (4.5). To investigate the effect of increase of backward delay, $T_{\mathrm{BW}}$, has been increased from $0.1s$ to $0.5s$, while $T_{\mathrm{FW}} = 0.2s$. The effect can be described as this: with the increase of backward delay time, the waiting time for the control increases

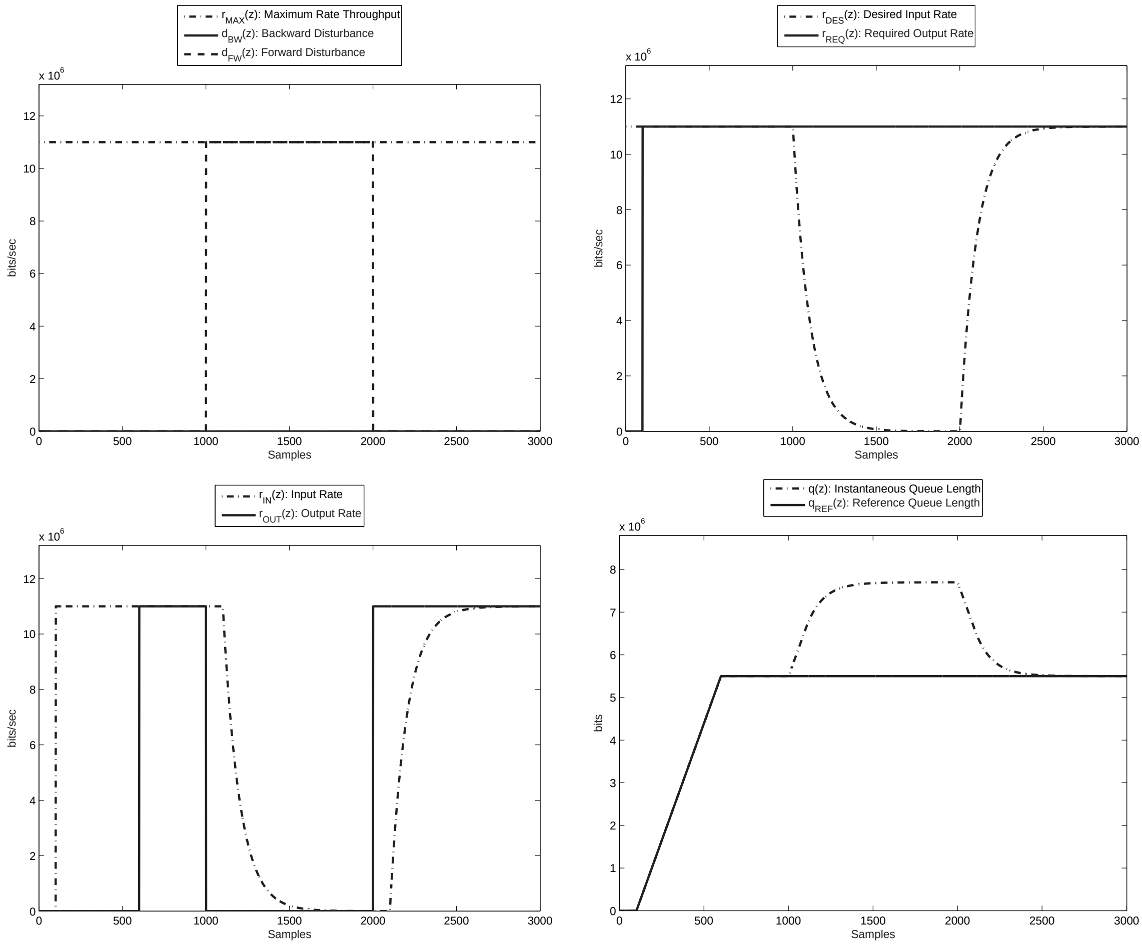


FIGURE 5.5: Simulation 5: Effect of forward delay, $T_{\text{FW}} = 0.5s$, $T_{\text{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

further for which the compensation becomes even bigger. Since Smith predictor feeds back this delay causing the system accumulated the time delay effects for the next compensation to be done. But due to the fact that $T_{\text{BW}} > T_{\text{FW}}$, the compensation becomes large enough to drive the rates negative and often overshoots take place since the pole is now outside the unit circle according the equation (4.5). However this instability problem can be practically handled using lower value limiter for the input and output rates so that negative values are not allowed. This will also allow the system to avoid overshoots since the error will reduce, requiring less compensation to be fed back to the desired rate, $r_{\text{DES}}(z)$. The instable system with overshoots due to higher backward loop delay is shown in Figure 5.6.

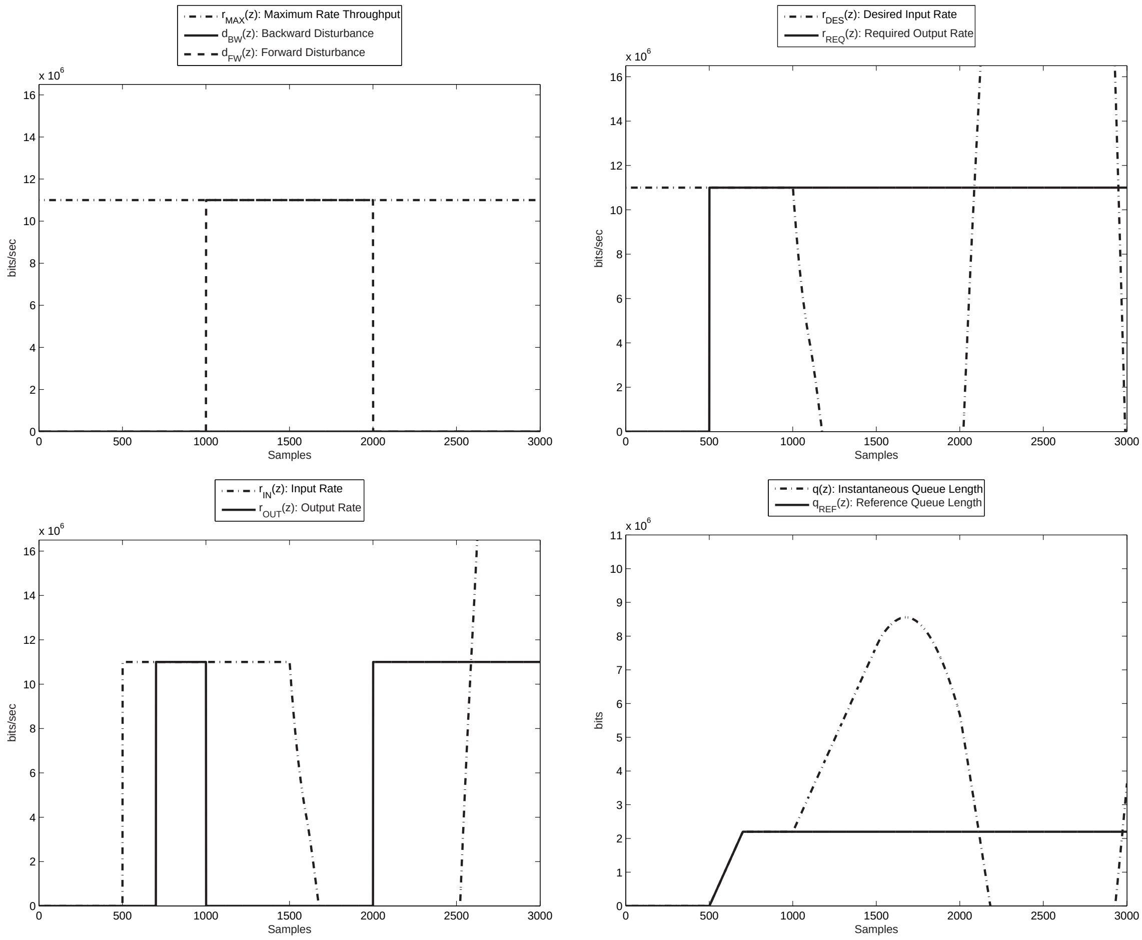


FIGURE 5.6: Simulation 6: Effect of backward delay, $T_{\mathrm{FW}} = 0.1s$, $T_{\mathrm{BW}} = 0.5s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

## 5.3 Simulation in WLAN

Based on the derivations in Sec. 4.1, the following parameter values are assumed for simulation in a WLAN scenario: $T_s = 0.001s$, $r_{\mathrm{MAX}}(z) = 11Mbps$, $T_{\mathrm{BW}} = 100T_s$, $T_{\mathrm{FW}} = 200T_s$, $T_{\mathrm{K}} = 100T_s$. Similarly like 5.1, the choice of $T_s$ is such that forward and backward delays can be effectively implemented using delay filters, as mentioned before. In order to allow iterative simulation, $r_{\mathrm{MAX}}(z)$, $d_{\mathrm{BW}}(z)$, $d_{\mathrm{FW}}(z)$, $r_{\mathrm{IN}}(z)$ and $r_{\mathrm{OUT}}(z)$ are provided with buffer lengths of 600 each for the delays which can be implemented by filters shown in Sec. 4.1. In all the following simulations, the initial queue length is assumed to be of zero length.

Using the scenario in the Figure 3.1, the simulation is performed using an OPNET (by MIL3) model of the filter based controller and plotted using MATLAB with exported data. Since control time and frame times are high level layer issues, in this simulation, these values are suitably chosen as integer multiple of sample time, i.e. control time, $T_c = 500T_s$ and frame time, $T_{fr} = 400T_s$. Nodes $A$, $C$, $G$ and $F$ are considered as the source nodes, while nodes $M$, $L$, $K$ and $J$ are considered as the destination nodes. Nodes $B$, $E$ and $I$ are the master nodes that assign the bandwidth among the cluster nodes, and nodes $D$ and $H$ act as bridges. To mimic the overall controller behaviour, each source and destination pair is considered to possess different maximum rate, $r_{\mathrm{MAX}}(z)$ and the nodes are considered to have a trajectory which is slower than the time required to update the topology information. The start and stop time for the connections and corresponding maximum rates are shown below

| **Connection** | **Start Time, s** | **Stop Time, s** | $r_{\mathrm{MAX}}(t)$**, Mbps** |
|---|---|---|---|
| $A \rightarrow M$ | 0.2 | 7.3 | 11.0 |
| $C \rightarrow L$ | 2.0 | 5.5 | 8.3 |
| $G \rightarrow K$ | 1.8 | 10 | 4.6 |
| $F \rightarrow J$ | 4.5 | 9.2 | 10.2 |

Table 5.1: Start and stop times for connections

In this simulation, only queue length of node $H$ is investigated since it is typically inflicted with more inbound and outbound requests during the transmissions. Also, time domain results are shown instead of z-domain, to illustrate the time-dependent results. As can be seen from Tab. 5.1, no transmission takes place up to $0.2s$ and as such the queue lengths increase linearly according to equation 4.3. Since from $t = 0.2s$ to $t = 1.8s$, only node $A$ is transmitting, $d_{\mathrm{BW}}(t) = 0Mbps$ and $d_{\mathrm{FW}}(t) = 0Mbps$. A random bandwidth sharing criterion for the master node has been considered to simulate congestion situation arbitrarily. From $t = 1.80s$ to $t = 2.0s$, nodes $A$ and $G$ start to transmit simultaneously as shown in Figure 5.7. During this interval, the controller copes to mitigate the disturbance through decreasing $r_{\mathrm{DES}}(t)$ and increasing $r_{\mathrm{REQ}}(t)$. This causes $q(t)$ to fall down with the reference queue length due to effect described in equation 4.11 and 4.2. Eventually $r_{\mathrm{IN}}(t)$ also falls down to lower value compared to $r_{\mathrm{OUT}}(t)$ during $T_{\mathrm{FW}} = 0.2s$ interval.

With the integrated scenario, the overall system has the following average queue length, $q_{\mathrm{AVG}}$ in bits and average throughput rate, $r_mathrmOUT_{AVG}$ in bps as shown in Table 5.2. These results are obtained from OPNET global link statistics and scaling the connection properties linearly according to their maximum throughput rates. It is evident that when forward disturbance persists, system

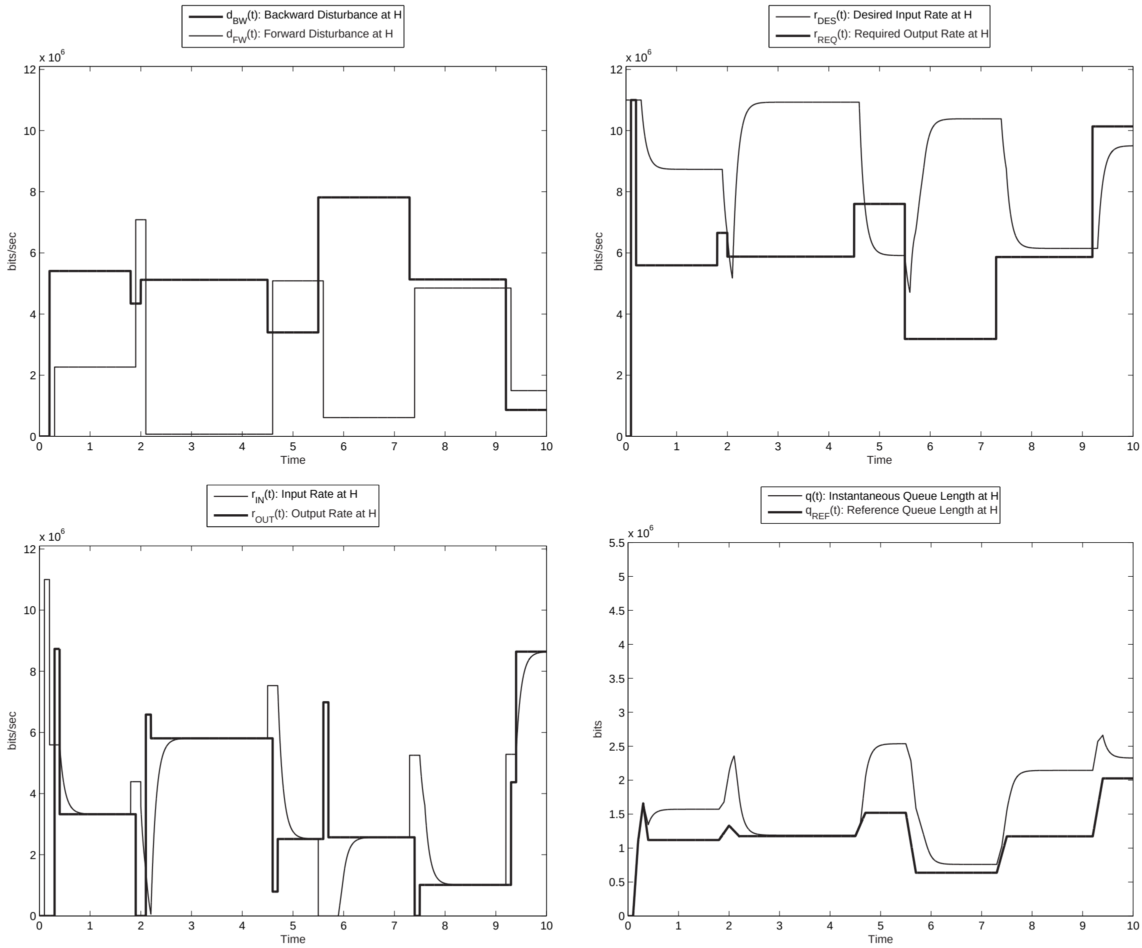


FIGURE 5.7: Simulation 7: Integrated WLAN Simulation, $T_{\mathrm{FW}} = 0.1s$, $T_{\mathrm{BW}} = 0.2s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

queue length exponentially increases and when a high backward disturbance appears in absence of forward disturbance, the queue length retains its IMC principles. The results in the table also validates the effectiveness of the proposed controller since it improves the data rate significantly, while maintains controllable queue length, which is not overflown. However, the only problem is that system has to scale its output and input rate at any node almost every sample time. But since the proposed controller is hardware based, it is possible for the controller to improve the scalability further. This is a strict requirement for networks, where the open system interconnect (OSI) layers are poorly defined, like in sensor networks or device networks etc.

| Connection | Connection Time, s | $q_{\text{AVG}}$, bits | $r_{\text{OUT}_{\text{AVG}}}$, bps |
|---|---|---|---|
| $A \rightarrow M$ | $0.2 \longleftrightarrow 7.3$ | $1.2718 \times 10^6$ | $3.5323 \times 10^6$ |
| $C \rightarrow L$ | $2.0 \longleftrightarrow 5.5$ | $0.7482 \times 10^6$ | $2.4519 \times 10^6$ |
| $G \rightarrow K$ | $1.8 \longleftrightarrow 10$ | $0.45381 \times 10^6$ | $1.93572 \times 10^6$ |
| $F \rightarrow J$ | $4.5 \longleftrightarrow 9.2$ | $1.3726 \times 10^6$ | $3.18268 \times 10^6$ |

TABLE 5.2: Average queue length and average throughput rate

## 5.4 Comparisons

The proposed filter-based controller is rather a lower level solution in modern network architecture, but surely improves on the kind of quick response that ad hoc networks may need. It can be easily implemented using modern DSP hardware. The only dilemma is that rates need to be adjusted almost on every sample instant. Also, for every connection, a bridge would need to have a dedicated filter-based controller, increasing the bulk of hardware. However, in [31], the continuous-time model for congestion control has been considered without any direction to whether it could be implemented as a hardware or software solution. Also, the proposed system considers the introduction of minimum value logic such that the input and output rates would never fall below zero. But in [31], this has not been done. This leads to impractical control when either $r_{\text{IN}}(t)$ or $r_{\text{OUT}}(t)$ falls below zero, due to compensation of $r_{\text{DES}}(t)$ through $r_{\text{DES}_{\text{ERR}}}(t)$. The comparisons are drawn in Figure 5.8.

To illustrate the differences, simulations are carried out with the proposed model and the model described in [31], whereby both are shown in the time-domain. Figure 5.8 shows the disturbances considered over a time duration of $10s$. With the introduction of the logic for the lower limit of $r_{\text{IN}}(t)$ and $r_{\text{OUT}}(t)$ as zero, the control objectives are still met and remains practical as shown in the Figure 5.8. Note that at $t = 3.5s$, $q(t)$ does not rise exponentially as the controller externally sets the lower value as $r_{\text{IN}_{\text{MIN}}}(t) = 0$ instead of negative value, while according to [31], the rates go negative, which is not realistic. The effect of negative rates on the queue lengths are shown in Figure 5.8 along with the realistic assumptions.

The proposed congestion control scheme is based on end-to-end congestion control. To illustrate the performance of proposed scheme, OPNET simulations results are shown in Table 5.3 with the following end-end-end transfer delays for the proposed scheme and the feedback based control scheme.

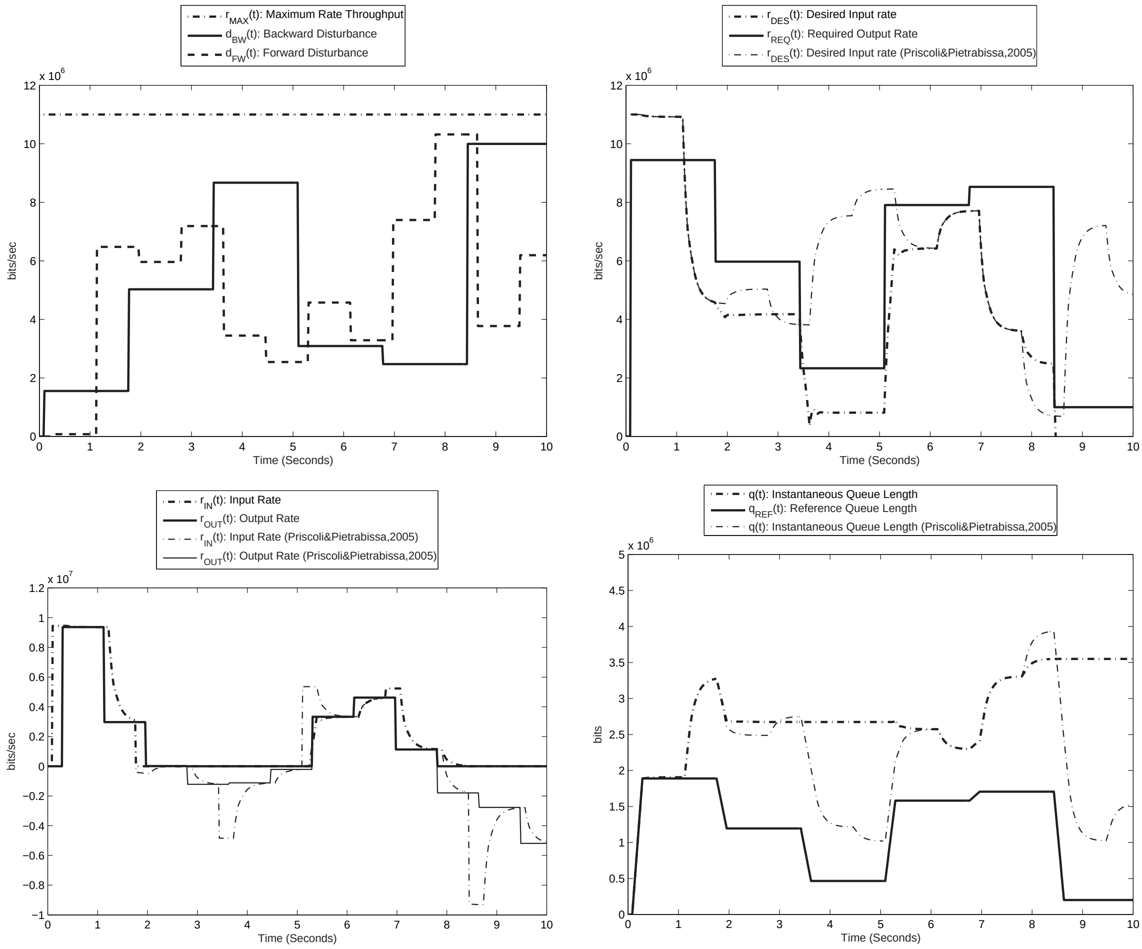


FIGURE 5.8: Simulation 8: Integrated Simulation for comparison , $T_{\mathrm{FW}} = 0.2s$, $T_{\mathrm{BW}} = 0.1s$: top left, maximum rate and disturbances, top right, desired and required rates, bottom left, input and output rates, bottom right, instantaneous and reference queue length.

| **Connection** | **Transfer delay with proposed scheme, s** | **Transfer delay with traditional feed-back based scheme, s** |
|---|---|---|
| $A \rightarrow M$ | 0.572 | 0.674 |
| $C \rightarrow L$ | 0.925 | 0.893 |
| $G \rightarrow K$ | 0.472 | 0.531 |
| $F \rightarrow J$ | 0.718 | 0.703 |

TABLE 5.3: Comparison of average transfer delays

## 5.5 Comments and Discussions

In this chapter, several simulations were carried out to demonstrate the system performance under varied circumstances. First, system performance was validated using MATLAB Simulink simulations and later an integrated WLAN scenario was simulated using OPNET discrete event simulator tool. Proposed model in this research considers certain advantages, as illustrated below:

1. In this proposed method, a realistic maximum buffer length has been considered as compared to [34], where the buffers are considered unrealistically large to contain all the receive traffic. The minimum buffer size that allows the proposed control system to work conforming with the control objectives has been defined in equation 3.18.
2. As shown in [31] and also in [30], the control scheme does not consider the fact that the input and output rates can go negative due simultaneous high backward and forward congestion. But in the proposed scheme strictly maintains the rates at or above zero, which is practical.
3. In networks sensor networks or device networks, ad hoc wireless LANs normally do not strictly maintain OSI layering model and as such this kind of model can be suitable for maintaining high data rate. It is particularly an advantage in today high speed networks, that adopt ad hoc mode and have bandwidth-on-demand system. Also since the implementation is proposed using high speed DSP hardware, the solution is fast and tunable at the same time.
4. The proposed scheme uses time division multiplexed request assignment cycle and as such, the time delays can be tuned towards stable system performance. Also, it is incumbent to design the minimum time delays for the system to achieve best results.
5. This scheme is an excellent candidate for congestion control in satellite networks, since the time delays are generally large and large queues can be implemented, too.

Despite these advantages, the system also has certain drawbacks. These are described below:

1. Since in the proposed scheme, the data rates need to be changed according to the time-delay control system, the rate adjustment can cause scalability problem. However, since this is hardware based, this scalability problem can be minimized using hardware level control on these rate using the control scheme described in this research.

2. Proposed scheme can be difficult to implement since it does not fit in the traditional OSI or TCP/IP model of network layering.

3. Due to added cost for DSP hardwares for each connection that a virtual access point has to deal with, the cost may increase. This is a trade off against all the advantages described above.

# Chapter 6

# Conclusions and Further Research

## 6.1 Conclusions

In this thesis, a congestion control scheme has been proposed from end-to-end control-theoretic paradigm with a DSP based implementation for multihop ad hoc wireless LANs with bandwidth-on-demand access. The proposed model in this research is novel in that it considers the controlling of congestion by means of processing the disturbance rate signals. The proposed model also improves on previous models, to ensure the control is realistic. Here, the basic control objectives are met while controlling congestion. In order to illustrate that the proposed system works as intended, a simple and generic WLAN scenario has been simulated. Since the proposed congestion control is real-time control, it avoids and controls congestion by making sure that the desired and required rates are regulated optimally according to the stated control objectives. Digital filter-based approach allows high-speed congestion control and a scalable hardware solution. The proposed congestion control scheme also uses combinatory stable assumptions, which is characteristic to WLANs. A more generic and quasi-stable assumptions are expected to lead to ground-breaking models for mobile ad hoc networks (MANETs). Recently research is being carried out for congestion control in MANETs with introduction of variable packet fragmentation and service-on-demand systems.

## 6.2 Future Directions

An ad hoc mode of operation for networked systems is presently being considered an attractive field of research and as such more speculations and observations are necessary for different applications using the same technology as described in this research. Next possible directions for ongoing/future research are outlined together with a summary of the challenges likely to be involved.

### 6.2.1 Congestion Control in Device Networks

Device networks support short-range wireless connections between devices and they are primarily more stable than present assumptions in terms of topology changes. Such networks are intended to replace inconvenient cabled connections with wireless connections. Thus, the need for cables and the corresponding connectors between cell phones, modems, headsets, computers, printers, projectors, network access points, and other such devices is eliminated. The interaction among the devices are less frequent and their network can also be set up on ad hoc basis. Also, master stations can be almost statically decided due to less change in topology and as such proposed congestion control scheme will fit easily. Also, lack of higher layers like in OSI network layer model can be an added advantage for the implementation. However, research opportunities lie in congestion control for a large device network, where balanced clustering is still a difficult problem [16].

### 6.2.2 Congestion Control in Sensor Networks

Senors have been in use for quite long time but wireless networked sensors are modern concepts with enormous potentials for research. Wireless sensor networks consist of small nodes with sensing, computation, and wireless networking capabilities, as such these networks represent the convergence of three important technologies. Sensor networks have enormous potential for both consumer and military applications. Military missions require sensors and other intelligence gathering mechanisms that can be placed close to their intended targets. The potential threat to these mechanisms is therefore quite high, so it follows that the technology used must be highly redundant and require as little human intervention as possible. Due to these attributes, ad hoc mode of operation for such network is an attractive solution but the fact that these networks perform and often exchange data in multiple path, make congestion an issue where intensive research opportunities remain.

One of the basic advantages of an ad hoc mode of sensor networks is that their network layers are very poorly defined and as such the proposed control scheme can be easily fit in. Unlike in WLANs, the master nodes in such network can be fixed because of the redundancy in the network. Using rather predictive delay and using adaptive Smith predictor, the present scheme can be configured to work for these networks, as well. However, the major challenge lies in prediction versus Smith predictor implementation and is being considered as an extension of this work. Such networks are being more used in today's feedback control, biological sensory networks, remote sensing and geographical information system etc.

### 6.2.3 Congestion Control in MANETs

Due to its wide range of scope for applications in today's fast growing communication technology, MANETs have attracted research interest for a long time. Initiated by Department of Defense (DoD) of the United States of America, it was first named as packet radio network. However, due to the limitations of the mobile nodes such as power and processing capability interests in this area were declined until quite lately. Recently, due to the development of high speed modern chip technology, faster adaptive power processing and low energy solution, MANETs have returned into mainstream research. Also, since in recent years demands of intercommunication using hand held devices without relying on fixed infrastructure such as base stations have been growing significantly, MANETs have become a core attention for research in modern wireless communication [28].

As much as MANETs offer outstanding possibilities, it also poses great challenge in congestion control since the topology change is even faster. Hence, combinatory stable assumptions made in this research does not hold any more. However, with the introduction of adaptive power control such assumptions can still be made. The only challenge that remains to be solved is that the scalability becomes even more of an non-trivial issue. This work can be generalized to fit into MANETs with the introduction of on-board power control and rate adjustment down to hardware level and is being considered as another direction for research.

### 6.2.4 Congestion Control for Service-on-Demand Systems

In the proposed scheme, the control of congestion is based on the bandwidth demand. However, with the advent of highly critical service oriented systems, the system performance is not affected by bandwidth only rather by a group of statistical and non-statistical variables. This leads to a classical problem of multi-dimensional control, where number of control variables both at the input and at the output side are more than one. In such system, the system can ensure a certain quality of service (QoS) is maintained as demanded while controlling the congestion and the queue length. Today's high speed satellite and other ad hoc networks use QoS parameters to indicate whether a service is acceptable or not. Presently, research works are being carried out to devise a novel service oriented technique to control the congestion.

# Appendix A

# Simulation with MATLAB Simulink

## A.1 Introduction to Simulink

Simulink is a software package for modeling, simulating, and analyzing dynamic systems. It supports linear and nonlinear systems, modeled in continuous time, sampled time, or a hybrid of the two. Systems can also be multirate, i.e., have different parts that are sampled or updated at different rates. It is a tool for model-based design and with Simulink, one can move beyond idealized linear models to explore more realistic nonlinear models, factoring in friction, air resistance, gear slippage, hard stops, and the other things that describe real-world phenomena. Simulink turns simple personal computers into a lab for modeling and analyzing systems that simply wouldn't be possible or practical otherwise, whether the behavior of an automotive clutch system, the flutter of an airplane wing, the dynamics of a predator-prey model, or the effect of the monetary supply on the economy. For modeling, MATLAB Simulink provides a graphical user interface (GUI) for building models as block diagrams, using click-and-drag mouse operations.

## A.2 Simulink Model

In this thesis, Simulink has been extensively used in several simulations to assess the system performance under varied circumstances. Figure A.1 shows the main simulation model that has been used in this research.

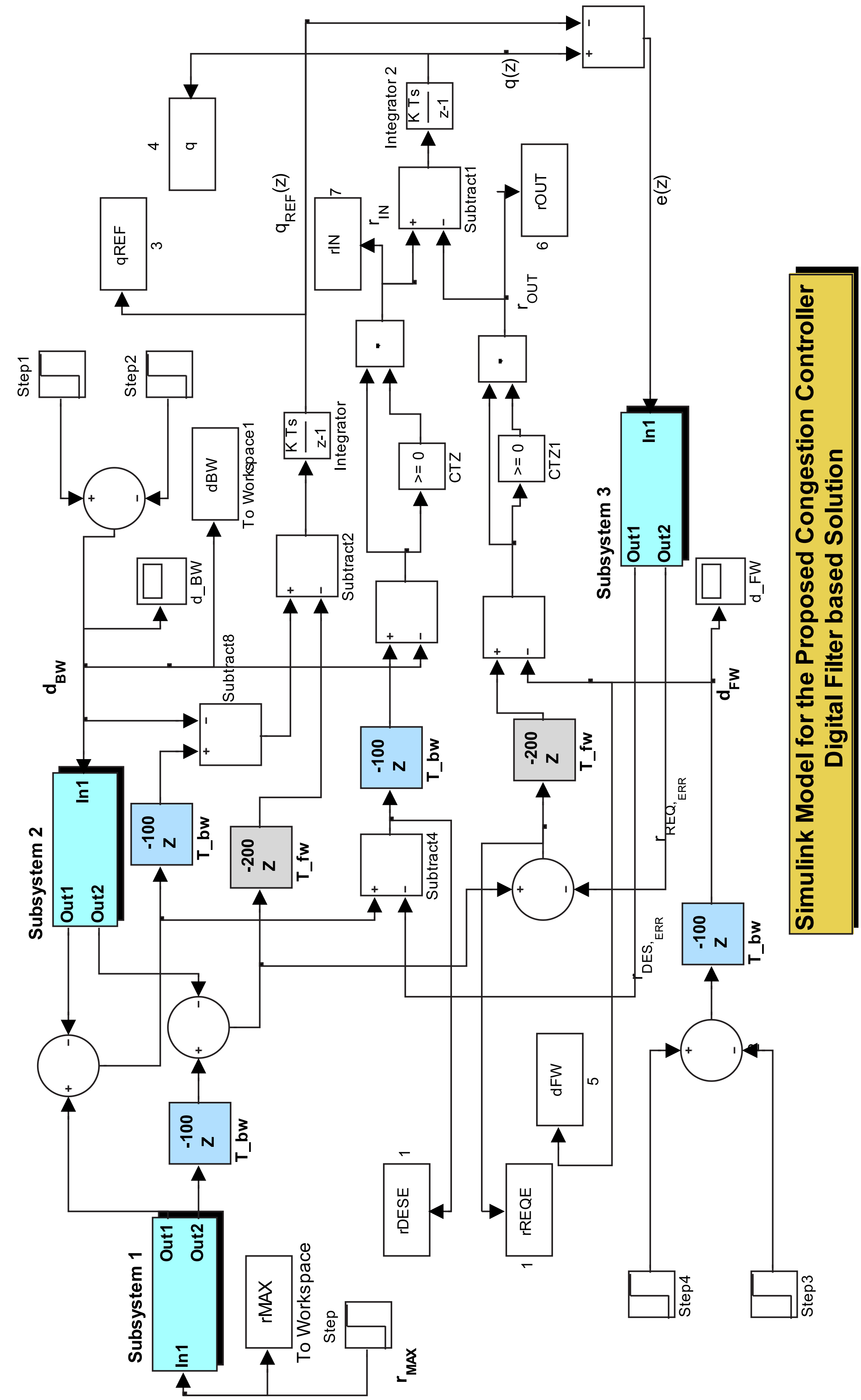


FIGURE A.1: Simulink model used for simulation

Subsystem 1 and Subsystem 2 are simple direct binary switching based on equation (4.3). Figure A.2 also shown the subsystem 3 which sets the error compensation for simulation.

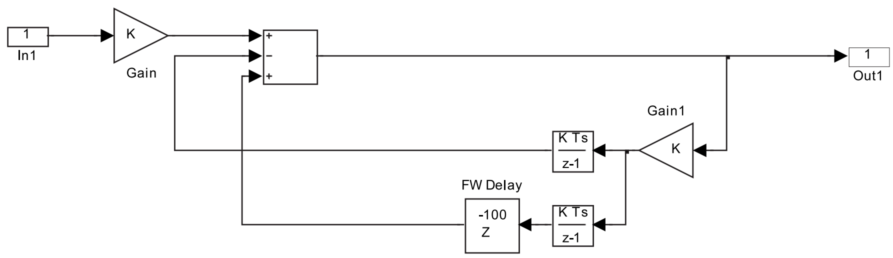


FIGURE A.2: Simulink subsystem 3 used in simulation

In the simulations, $T_f w$ is used for forward delay, which is set to $200T_s$ and $T_b w$ is used for backward delay, which is set to $100T_s$. The $CTZ$ boxes are simulink compare to zero boxes. The system is implemented with 600 buffer length. All other symbols have their usual meanings described in Chapter 3 and 4.

# Appendix B

# Simulation with OPNET Model

## B.1 Introduction to OPNET

OPNET is a a commercial tool by MIL3, Inc., OPNET (Optimized Network Engineering Tools) is an engineering system capable of simulating large communication networks with detailed protocol modeling and performance analysis. It's features include graphical specification of models, a dynamic, event-scheduled Simulation Kernel, integrated data analysis tools and hierarchical, object based modeling. It is a network simulation tool that allows the definition of a network topology, the nodes, and the links that go towards making up a network. The processes that may happen in a particular node can be user defined, as can the properties of the transmission links. A simulation can then be executed, and the results analyzed for any network element in the simulated network [7].

The key features of OPNET are that, it provides powerful tools that assist the user in the design phase of a modeling and simulation project, i.e., the building of models, the execution of a simulation and the analysis of the output data. OPNET employs a hierarchical structure to modeling, that is, each level of the hierarchy describes different aspects of the complete model being simulated. It has a detailed library of models that provide support for existing protocols and allow researchers and developers to either modify these existing models or develop new models of their own. Furthermore, OPNET models can be compiled into executable code. An executable discrete-event simulation can be debugged or simply executed, resulting in output data. OPNET has three main types of tools - the Model Development tool, the Simulation Execution tool and the Results

Analysis tool. These three types of tools are used together to model, simulate and analyze a network.

## B.2 OPNET Model

In this thesis, a rather simple WLAN scenario has been simulated using OPNET discrete event simulator, version 8.1. The WLAN is designed according to figure-fig:adhoc in Chapter 3. The detailed settings of the OPNET model is available at author's discretion. Figure B.1 shows the simulation model that has been used to investigate the system performance in real WLAN environment.

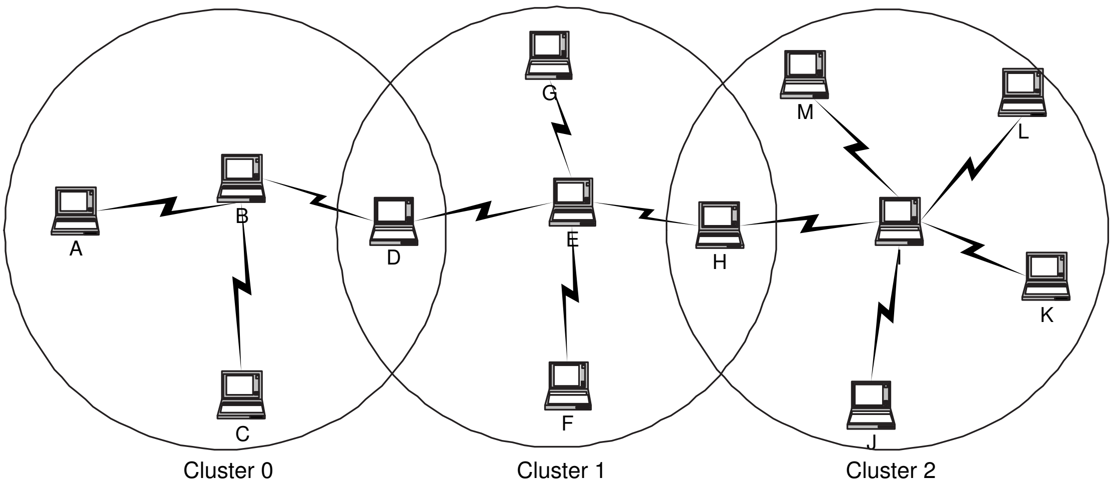


FIGURE B.1: OPNET model used for simulation

Apart from the parameters shown in Table 5.1, parameters and their values are used for the simulation of a realistic WLAN scneario are shown in Table B.1.

| **Symbol** | **Parameter** | **Value** |
|---|---|---|
| $T_{\text{frame}}$ | Frame length | $0.081s$ |
| $T_{\text{feedback}}$ | Feedback Delay | 7 frames |
| $T_{\text{K}}$ | Buffer filling allowance time | $6.7s$ |
| $C_0$ | Capacity of Cluster 1 | 80 packets/frame |
| $C_1$ | Capacity of Cluster 2 | 80 packets/frame |
| $C_2$ | Capacity of Cluster 3 | 60 packets/frame |
| $T_{\text{C}}$ | Control frame size | 5 frames |

TABLE B.1: Parameter values used in OPNET simulation

# Bibliography


[1] G. Açar and C. Rosenburg. “Weighted fair bandwidth-on-demand (WFBoD) for geostationary satellite networks with on-board processing”. *Computer Networks*, 39(1):5–20, 2002.

[2] E. Altman, T. Başar, and T. Srikant. “Congestion control as a stochastic control problem with action delays”. *Automatica*, vol.35(No.12), 1999. pp.1937–1950.

[3] S. Athuralia, V. Li, S. Low, and Q. Yin. “REM: active queue management”. *IEEE Networking*, 15(3):pp.48–53, 2001.

[4] S. Basagni, Turgut.D., and S. Das. “Mobility-adaptive protocols for managing large ad hoc networks”. In *Proceedings of IEEE International Conference*, pages 1539–1543, June 2001.

[5] R. Braatz. *The Control Handbook*. CRC Press, W.S.Levine,Ed., Boca Raton, FL, 215–224 edition, 1996.

[6] L. Brakmo and L. Peterson. “TCP Vegas end-to-end congestion avoidance on a global Internet”. *IEEE Journal on Selected Areas in Communications*, Vol.13(No.8):pp.1465–4, 1995.

[7] X. Chang. Network Simulation using OPNET. In *Proceedings of the 1999 Simulations Conference*, volume 5, pages pp.95–104, 1999.

[8] C. Chicoix, J. Pedoussat, and N. Giambiasi. “An accurate time delay model for large digital network simulation”. *Proceedings of the 13th Conference on Design Automation*, pages 54–60, 1976. San Fransisco, California, USA.

[9] L. Chisci, Fantacci, and T. Pecorella. “Predictive bandwidth control for GEO satellite networks”. *Proceedings of International Conference on Communications, ICC'04*, pages 3958–3962, 2004.

[10] P. Dyson. *Dictionary of Networking*. Sybex Inc., 1151 Marine Village Parkaway, Alameda, CA 94501-1044, USA, 3rd edition, 2003. ISBN: 0-7821-2461-5.

[11] A. Ephremides, J. Wieselthier, and D. Baker. "A design concept for reliable mobile radio networks with frequency hopping signalling". In *Proceedings of IEEE*, pages 56–73, Jan. 1987.

[12] M. Fisk and W. Feng. "Dyanmic Right Sizing in TCP". In *Proc. of the Los Alamos Computer Science Institute Symposium*, LA-UR, 01-5460, Oct. 2001.

[13] S. Floyd and V. Jacobson. "Random early detection gateways for congestion avoidance". *IEEE/ACM Transaction on Networking*, Vol.1(No.4):pp.397–413, January 1993.

[14] A. Freedman, editor. *Computer Desktop Encyclopedia*. AMACOM, 2nd bk and cdr edition, 1999. ISBN: 0814479855.

[15] J. Geier and J. Geier. *Wireless Networking Handbook*. New Riders Publising, Indianapolis, 1996.

[16] A. Goldsmith. *Wireless Communcaitions*. Cambridge University Press, 1st edition, August 2005.

[17] K. Gu, V. Kharitonov, and J. Chen. *Stability and Robust Stability of Time-delay Systems*. Birkhauser, Boston, USA, 2003.

[18] P. Ioannou. *The Control Handbook*, chapter Model Reference Adaptive Control, pages pp.847–858. CRC Press, 1996.

[19] V. Jacobson. "Congestion Control and Avoidance". *Computer Communications Review*, vol.18(4):314–329, Aug. 1988.

[20] K. Jacobsson, H. Håkan, N. Möller, and K. Johansson. "Round Trip Time Estimation in Communication Networks Using Adaptive Kalman Filtering". In *Proceedings of IEEE Networks*, pages 156–160, 2004.

[21] R. Karri and Mishra.P. "Modeling energy efficient secure wireless networks using network simulation". In *Proceedings of IEEE Internation Conference of Communication*, pages 61–65, May 2003.

[22] S. Low, L. Peterson, and L. Wang. "Understanding Vegas: a duality model". *Journal of ACM*, Vol.49(No.2):pp.207–235, 2002.

[23] S. Mascolo. "Congestion Control in high-speed communication networks". *Automatica*, Vol.35(No.12):pp.1921–1935, 1999.

[24] S. Mascolo. "Modeling the internet congestion control using Smith controller with input shaping". *Control Engineering Practice*, Vol.13(2005):pp.312–319, 2005.

[25] C. Murthy and S. Manoj, B. *Ad Hoc Wireless Networks: Architectures and Protocols*. Prentice Hall Communications and Emerging Techno, May 24 2004.

[26] J. Nagle. *Congestion Control in IP/TCP Internetworks*. ARPANET Working Group Requests for Comment, DDN Network Information Center, RFC-896, SRI International, Menlo Park, CA, Jan. 1984.

[27] S. Niculescu. *Delay effects of stability: A robust control approach*, volume vol.269. Springer-Verlag Heidelberg, 1995.

[28] P. Pham. *Congestion Avoidance using Multipath Routing and Power Control in Mobile ad hoc Networks*. PhD thesis, University of South Australia, 2002.

[29] F. D. Priscoli and A. Isidori. "A control engineering approach to integrated congestion control and scheduling in local area networks". *Control Engineering Practice*, vol.13(2005):pp.541–558, 2005.

[30] F. D. Priscoli and A. Pietrabissa. "Design of a capacity-on-demand (BoD) protocol for satellite networks modelled as time-delay systems". *Automatica*, 40(5):729–741, 2004.

[31] F. D. Priscoli and A. Pietrabissa. "Hop-by-Hop Congestion Control for Wireless LAN: A Model-Based Control Approach". *International Journal of Control*, 2005. to appear.

[32] M. Pursley. "The role of spread spectrum in packet radio networks". *IEEE Proceedings*, pages pp.116–134, January 1987.

[33] P. Quet, B. Ataşlar, A. Iftar, H. Özbay, S. Kalayanaraman, and T. Kang. "Rate based flow controllers for communication networks in the presence of uncertain time-varying multiple time-delays". *Automatica*, Vol.38(No.6):pp.917–928, 2002.

[34] G. Ramamurthy and B. Sengupta. "A Predictive Hop-by-Hop Congestion Control Policy for High-Speed Networks". *In Proceedings of INFOCOM'93*, pages 1033–1041, 1993.

[35] R. Ramanathan and Rosales-Hein. "Topology control of multihop wireless networks using transmit power adjustment". In *Proceedings of IEEE INFOCOM'00*, pages 404–413, March 2000.

[36] R. Shafik, S. Weiss, and E. Rogers. "A DSP-based approach to congestion control for multihop ad hoc WLANs with BoD access". In *2nd IEE-EURASIP DSPEnabledRadio Conference'05*, pages pp.31/1–31/9, Southampton, SO17 1BJ, September 2005.

[37] O. Smith. "A controller to overcome dead time". *ISA Journal*, Vol.6:28–33, 1959.

[38] W. Stalling. *High Speed Networks: TCP/IP and ATM Design Principles*. Prentice Hall, San Francisco, 1998.

[39] W. Stevens. *TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms*. Network Working Group Requests for Comment, RFC-2001, Category Standards Track, NOAO, Jan. 1997.

[40] L. Tan, A. Pugh, and M. Yin. "Rate Based Congestion Control in ATM switching networks using a recursive digital filter". *Control Engineering Practice*, Vol.11(2003):pp.1171–1181, February 2003.

[41] F. Tobagi. "Modeling and performance analysis of multihop packet radio networks". *IEEE Proceedings*, pages 135–155, January 1999.

[42] E. Walter, editor. *Cambridge Advanced Learner's Dictionary*. Cambridge University Press, 2nd edition, April 15 2005. ISBN: 0521604982.

[43] Y. Yi and S. Shakkottai. "Hop-by-hop congestion control over a wireless multihop network". In *IEEE INFOCOM'04*, pages pp.1214–1224, July 2004.