A DSP-BASED APPROACH TO CONGESTION CONTROL FOR MULTIHOP AD HOC WLANS WITH BOD ACCESS

Rishad Ahmed Shafik, Stephan Weiss, and Eric Rogers

School of ECS, University of Southampton, UK

Keywords: Bandwidth-on-demand, Ad Hoc WLAN, Model Reference Controller, Smith Predictor, Digital Filter.

Abstract

In this paper, a DSP-based solution has been proposed 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 against. 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 has been used in the closed-loop error controller to compensate for delays that could cause instability. 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. Simulations illustrate the system performance of the filter-based controller in a simple WLAN model. Finally, a brief comparative study has been made with some of the recently proposed models that perform similar control tasks.

1 Introduction

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 [5]. These LANs can be operated with different modes of access and routing systems; in this paper we consider multihop routing of data packets and bandwidth-on-demand (BoD) access mechanism. Multihop routing involves routing of packets through intermediate nodes in multiple hops in order to transmit data from the source to the destination. Essentially, in ad hoc networks, this is possible when some of the intermediate nodes act as bridges. 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 [9].

In this paper, the BoD controller is termed the "master node" and each node is assigned a particular time-slot in the request assignment cycle, after which the BoD controller has complete knowledge of which nodes want to transmit and at what rate. Similar BoD protocols for satellite networks have been developed by Açar and Rosenburg, [1] and later used in [8], [4] and [9]. In this proposed system, "congestion control" refers to controlling of queue lengths such that the packets are not dropped due to overflowing buffers. 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 term "cluster" refers to the group of wireless nodes that can access each other directly by a single hop and "bridge" refers to the dedicated wireless node that is visible (accessible) by nodes of two or more neighbouring clusters. 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 the following sections, a disretized and properly approximated model is then developed towards a digital-filter based solution for control of congestion. This can then be used as a means of continual control of congestion rather than on-demand control and mitigates the scalability problem to some extent. Filter models have been previously used in [12] to control the congestion of a ATM switching network, but 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.

2 System Model

In order to derive the system model, a basic ad hoc wireless network with BoD access is studied first. As shown in Fig. 2.1, 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 Fig. 2.1 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 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.

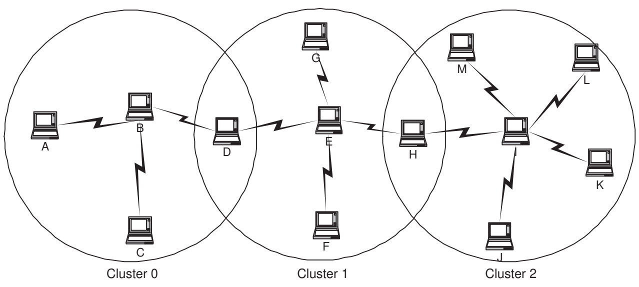

Fig. 2.1. A typical ad hoc wireless LAN

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. 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, bandwith requests from nodes in the cluster and 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 backward delay and will be denoted as T_{BW}. 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_{FW}. If the requested ouput rate, $r_{\mathrm{REQ}}(t)$, is higher than the allowed output rate, $r_{\mathrm{OUT}}(t)$, an additive forward disturbance takes place, which is denoted by $d_{\mathrm{FW}}(t)$. 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 $r_{\mathrm{OUT}}(t)$ allowed by the bridge controller, which is decided based on the disturbances and maximum throughput rate. Node H requests the required output rate, $r_{\mathrm{REQ}}(t)$, at which it wants to transmit.

6. Master node I allows the rate of $r_{\mathrm{OUT}}(t)$ as in step 4. At this moment, node H despatches the data directly to the node M of cluster 2 at a rate of $r_{\mathrm{OUT}}(t)$. 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, $r_{\mathrm{IN}}(t)$, is higher than the actual output rate, $r_{\mathrm{OUT}}(t)$, 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, $q_{\mathrm{REF}}(t)$, will be first deduced from the maximum data rate, $r_{\mathrm{MAX}}(t)$, and backward disturbance, $d_{\mathrm{BW}}(t)$, 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, $q_{\mathrm{REF}}(t)$. The overall system is intended to meet the following objectives:

1. The queue lenghts must be limited to maximum buffer size of S, with a lower limit of 0, i.e.

$$0 \leq q(t) \leq S \quad . \tag{2.1}$$

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, $r_{\mathrm{DES}}(t)$ and $r_{\mathrm{REQ}}(t)$ respectively, are driven back to the maximum data rate, $r_{\mathrm{MAX}}(t)$, 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 aymptotically stable independently of forward or backward disturbance.

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 [9] and later a digital filter-based solution is presented with necessary modifications.

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, $r_{\mathrm{IN}}(s)$, is higher than the output rate, $r_{\mathrm{OUT}}(s)$, i.e.

$$q(s) = \frac{1}{s}\left[r_{\mathrm{IN}}(s) - r_{\mathrm{OUT}}(s)\right] \quad . \tag{2.2}$$

In order to make sure that the objective 1 is met, the necessary condition is $\frac{1}{s}(r_{\mathrm{IN}}(s) - r_{\mathrm{OUT}}(s)) \leq S$, where S is the maximum buffer length. In [10] 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 2.18.

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

$$
\begin{aligned}
d_{\mathrm{FW}}(s) &= r_{\mathrm{REQ}}(s) - r_{\mathrm{OUT}}(s) \quad , &(2.3)\\
d_{\mathrm{BW}}(s) &= r_{\mathrm{DES}}(s) - r_{\mathrm{IN}}(s) \quad , &(2.4)
\end{aligned}
$$

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

$$
\begin{aligned}
0 \leq d_{\mathrm{FW}}(s) \leq r_{\mathrm{REQ}}(s) \leq r_{\mathrm{MAX}}(s) \quad , &\qquad (2.5)\\
0 \leq d_{\mathrm{BW}}(s) \leq r_{\mathrm{DES}}(s) \leq r_{\mathrm{MAX}}(s) \quad . &\qquad (2.6)
\end{aligned}
$$

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 2.2. 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,

$$r_{\mathrm{DES}_{\mathrm{REF}}}(s) = r_{\mathrm{MAX}}(s) \quad . \qquad (2.7)$$

From equations 2.5 and 2.6, 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 2.6 as $r_{\mathrm{MAX}}(s)$, the minimum value of $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ is 0. Thus, with the given contraint in equation 2.6, $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$ is given by

$$r_{\mathrm{REQ}_{\mathrm{REF}}}(s) = r_{\mathrm{MAX}}(s)e^{-sT_{\mathrm{BW}}} - d_{\mathrm{BW}}(s) \qquad (2.8)$$

where the delay term, $e^{-sT_{\mathrm{BW}}}$, is due to backward delay, T_{BW}, in the time-delay model caused by the master node in the backward acknowledgement process. Equation 2.8 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 [2].

Equations 2.7 and 2.8 show that the required reference output rate, $r_{\mathrm{REQ}_{\mathrm{REF}}}(s)$, can be redefined in terms of the desired reference output rate, $r_{\mathrm{DES}_{\mathrm{REF}}}(s)$,

$$r_{\mathrm{REQ}_{\mathrm{REF}}}(s) = r_{\mathrm{DES}}(s)e^{-sT_{\mathrm{BW}}} - d_{\mathrm{BW}}(s) \quad . \qquad (2.9)$$

During the time interval between requesting an output rate and acknowledgement 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 modelled by

$$q_{\mathrm{REF}}(s) = \frac{1}{s}\left[r_{\mathrm{DES}_{\mathrm{REF}}}(s) - r_{\mathrm{REQ}_{\mathrm{REF}}}(s)e^{-sT_{\mathrm{FW}}}\right]. \qquad (2.10)$$

Combining equations 2.7, 2.8 and 2.10, $q_{\mathrm{REF}}(s)$ can further be expressed in terms of inputs $r_{\mathrm{MAX}}(s)$ and $d_{\mathrm{BW}}(s)$ as

$$
\begin{aligned}
q_{\mathrm{REF}}(s) &= \frac{1}{s}[r_{\mathrm{MAX}}(s)(1 - e^{-s(T_{\mathrm{BW}}+T_{\mathrm{FW}})})\\
&\quad + d_{\mathrm{BW}}(s)e^{-sT_{\mathrm{FW}}}] \quad . \qquad (2.11)
\end{aligned}
$$

The delays in equation 2.11 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 [6]. 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.

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. 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 2.3 and 2.4, 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

$$e(s) = q(s) - q_{\mathrm{REF}}(s) \quad . \qquad (2.12)$$

Since the forward disturbance happens after $r_{\mathrm{OUT}}(s)$ has been announced by the master node after T_{FW}, the compensation of d_{FW} during this time requires that the error controller waits for the "dead time", T_{BW}, for the next error feedback to be available. This can cause instability and as such, control of $r_{\mathrm{IN}}(s)$ and $r_{\mathrm{OUT}}(s)$ by means of $e(s)$ must consider the effect of "dead time" according to Smith's principle [11]. Hence, transfer function between $e(s)$ and $d_{\mathrm{FW}}(s)$ can be given by

$$\frac{e(s)}{d_{\mathrm{FW}}(s)} = \frac{1 - e^{-sT_{\mathrm{BW}}}}{s} + \frac{e^{-sT_{\mathrm{BW}}}}{s + \frac{1}{T_K}}. \qquad (2.13)$$

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

$$T_K = \frac{S}{r_{\mathrm{MAX}}} - (T_{\mathrm{FW}} + T_{\mathrm{BW}}) \quad . \tag{2.14}$$

Since both terms in the sum in equation 2.13 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 2.13 also establishes the limiting values of $e(s)$. Taking the inverse Laplace transform of equation 2.13 yields

$$\begin{aligned} e(t) &= \int_{\mathrm{t-T_{BW}}}^{t} d_{\mathrm{FW}}(\tau)d\tau \\ &+ \int_0^{t-T_{\mathrm{BW}}} e^{-\frac{\tau}{T_{\mathrm{K}}}} d_{\mathrm{FW}}(t - T_{\mathrm{BW}} - \tau)d\tau. \end{aligned} \tag{2.15}$$

Considering $T_K > 0$, $d_{\mathrm{FW}}(t) = 0$ for $-(T_{\mathrm{BW}} + T_{\mathrm{FW}}) \leq t \leq 0$ and $0 \leq d_{\mathrm{FW}}(t) \leq r_{\mathrm{MAX}}$ for $t >= 0$ in equation 2.15 yields

$$\begin{aligned} 0 \leq e(t) &\leq r_{\mathrm{MAX}} \left\{ \int_{\mathrm{t-T_{BW}}}^{t} d\tau + \int_0^{t-T_{\mathrm{BW}}} e^{-\frac{\tau}{T_K}} d\tau \right\} \\ &\leq r_{\mathrm{MAX}} \left(T_{\mathrm{BW}} + T_K\right), \forall t \geq 0 \quad . \end{aligned} \tag{2.16}$$

Equation 2.16 is essentially limiting the queue error, $e(s)$, that is set by the controller and can be verified. In equation 2.9, when $d_{\mathrm{BW}}(s) = 0$, $r_{\mathrm{REQ_{REF}}}(s)$ is asymptotically driven to the value of $r_{\mathrm{DES_{REF}}}(s)$ and $q_{\mathrm{REF}}(s) = 0$. But when $d_{\mathrm{BW}}(s) \neq 0$, $q_{\mathrm{REF}}(s) > 0$ according to equation 2.10. Thus, the minimum error takes place when $q(s)$ and $q_{\mathrm{REF}}(s)$ are same, i.e.

$$e(s)_{\mathrm{MIN}} = 0 \quad . \tag{2.17}$$

On the other hand, from equation 2.12, it is evident that $e(s)$ is maximum when $q(s)$ is maximum and q_{REF} is minimum. According to control objective 1 and equation 2.14, the upper limit of $q(s)$ is given by

$$\begin{aligned} q_{\mathrm{MAX}}(s) &= r_{\mathrm{MAX}}(s) \left[e^{-sT_K} + e^{-sT_{\mathrm{BW}}} + e^{-sT_{\mathrm{FW}}}\right] \\ &= S \quad . \end{aligned} \tag{2.18}$$

From equation 2.10, $q_{\mathrm{REF}}(s)$ is minimum when $r_{\mathrm{REQ_{REF}}}(s)$ is maximum as shown in the inequality 2.5. The maximum value of $r_{\mathrm{REQ_{REF}}}(s)$ from equation 2.8 is $r_{\mathrm{MAX}}(s)$. Thus, $q_{\mathrm{REF_{MIN}}}(s)$ and the maximum error, $e_{\mathrm{MAX}}(s)$ can be shown as

$$\begin{aligned} q_{\mathrm{REF_{MIN}}}(s) &= \frac{1}{s} r_{\mathrm{MAX}}(s)[1 - e^{-sT_{\mathrm{FW}}}] \\ &= r_{\mathrm{MAX}}(s) e^{-sT_{\mathrm{FW}}} \quad , \end{aligned} \tag{2.19}$$

$$\begin{aligned} e_{\mathrm{MAX}}(s) &= q_{\mathrm{MAX}}(s) - q_{\mathrm{REF_{MIN}}}(s) \\ &= r_{\mathrm{MAX}}(s) \left[e^{-sT_K} + e^{-sT_{\mathrm{BW}}}\right]. \end{aligned} \tag{2.20}$$

In the time-domain, equations 2.17 and 2.20 can be combined as the following inequality

$$0 \leq e(t) \leq r_{\mathrm{MAX}}(T_K + T_{\mathrm{BW}}) \quad . \tag{2.21}$$

Inequality 2.21 is a direct result of control objectives 1 and 3. With known limits set in inequality 2.21, 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, $r_{\mathrm{DES_{ERR}}}(s)$ and required input rate error, $r_{\mathrm{REQ_{ERR}}}(s)$ as

$$r_{\mathrm{DES}}(s) = r_{\mathrm{DES_{REF}}}(s) - r_{\mathrm{DES_{ERR}}}(s), \tag{2.22}$$

$$r_{\mathrm{REQ}}(s) = r_{\mathrm{REQ_{REF}}}(s) - r_{\mathrm{REQ_{ERR}}}(s). \tag{2.23}$$

As specfied at the beginning of Sec. 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_{BW} by Smith's principle [11], we have the following transfer functions for the error rates,

$$\frac{r_{\mathrm{DES_{ERR}}}(s)}{e(s)} = \frac{s\frac{1}{T_K}}{s + \frac{1}{T_K}(1 - e^{-sT_{\mathrm{BW}}})}, \tag{2.24}$$

$$\frac{r_{\mathrm{REQ_{ERR}}}(s)}{e(s)} = 0. \tag{2.25}$$

Multiplying the transfer functions in equations 2.24 and 2.13, the following Laplace and time domain transfer functions can be shown,

$$\frac{r_{\mathrm{DES_{ERR}}}(s)}{d_{\mathrm{FW}}(s)} = \frac{\frac{1}{T_K}}{s + \frac{1}{T_K}}, \tag{2.26}$$

$$r_{\mathrm{DES_{ERR}}}(t) = \frac{1}{T_K} \int_0^t e^{-\frac{\tau}{T_K}} d_{\mathrm{FW}}(t - \tau)d\tau. \tag{2.27}$$

From equation 2.27 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 2.26 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

$$0 \leq r_{\mathrm{DES_{ERR}}}(t) \leq r_{\mathrm{MAX}}(t) \quad . \tag{2.28}$$

On the other hand, since $r_{\mathrm{REQ_{ERR}}}(t) = 0$ according to equation 2.25, 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 2.5 as

$$0 \le r_{\mathrm{REQ}}(s) \le r_{\mathrm{MAX}}(s) \quad . \tag{2.29}$$

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 [3].

3 Digital-filter based Design

In order to deduce a digital-filter based solution, the overall system derived in Sec. 2 is sampled with a sampling period, T_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_{BW} and T_{FW}, can be considered as integer multiples of T_s, $T_{\mathrm{BW}} = \alpha T_s$ and $T_{\mathrm{FW}} = \beta T_s$, where α and β are non-negative integers. Also, since $T_K >> 2T_s$ and $T_K = \gamma T_s$, where γ is also non-negative integer, from Nyquist theorem, it follows that the discertization will mimic the continuous time system, without any distortion or loss [7]. Following the continuous-time system, shown in Sec. 2.1 and in Sec. 2.2, and using the z-transform, the digital filters will be derived from their continuous counterparts in the following.

3.1 Reference Filter

Based on the reference controller in Sec. 2.1, a reference filter is found in this section using simplifications and transformations, as required. Based on the desired and required reference rate, $r_{\mathrm{DES_{REF}}}(s)$ and $r_{\mathrm{REQ_{REF}}}(s)$, shown in equations 2.7 and 2.8, tranformed and redefined rates in z-domain are given by

$$\begin{aligned} r_{\mathrm{DES_{REF}}}(z) &= r_{\mathrm{MAX}}(z) \quad , & (3.1)\\ r_{\mathrm{REQ_{REF}}}(z) &= r_{\mathrm{MAX}}(z) z^{-\alpha} - d_{\mathrm{BW}}(z) \quad . & (3.2)\end{aligned}$$

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 2.11, i.e.

$$\begin{aligned} q_{\mathrm{REF}}(z) &= \frac{1 - z^{-(\alpha+\beta)}}{z-1} r_{\mathrm{MAX}}(z) T_s \\ &+ \frac{z^{-\beta}}{z-1} d_{\mathrm{BW}}(z) T_s \quad . \end{aligned} \tag{3.3}$$

The expression for the reference queue length in equation 3.3 can be simplified as

$$\begin{aligned} q_{\mathrm{REF}}(z) &= \underbrace{\frac{z^{-1} - z^{-(\alpha+\beta+1)}}{1 - z^{-1}}}_{\text{IIR Filter}} r_{\mathrm{MAX}}(z) T_s \\ &+ \underbrace{\frac{z^{-(\beta+1)}}{1 - z^{-1}}}_{\text{IIR Filter}} d_{\mathrm{BW}}(z) T_s \quad . \end{aligned} \tag{3.4}$$

Equation 3.4 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 3.4, 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 $(\alpha + \beta + 1)$.

3.2 Error Filter

As seen in Sec. 2.2, based on the error between instantaneous queue length and $q_{\mathrm{REF}}(z)$, and found in Sec. 3.1, the error filter must generate the compensation for $d_{\mathrm{FW}}(z)$ as $r_{\mathrm{DES_{ERR}}}(z)$ and $r_{\mathrm{REQ_{ERR}}}(z)$. From quations 2.24 and 2.25, $r_{\mathrm{DES_{ERR}}}(z)$ and $r_{\mathrm{REQ_{ERR}}}(z)$ can be expressed as

$$\begin{aligned} \frac{r_{\mathrm{DES_{ERR}}}(z)}{e(z)} &= \frac{\frac{1}{\gamma T s}}{1 + \frac{1}{\gamma(z-1)}(1 - z^{-\alpha})} \\ &= \underbrace{\frac{\frac{1}{T_s}(1 - z^{-1})}{\gamma + (1-\gamma) z^{-1} - z^{-(\alpha+1)}}}_{\text{IIR Filter}} & (3.5) \\ \frac{r_{\mathrm{REQ_{ERR}}}(z)}{e(z)} &= 0. & (3.6) \end{aligned}$$

where $e(z)$ is the z-domain queue error and can be expressed from equations 2.12 as $e(z) = q(z) - q_{\mathrm{REF}}(z)$. The transfer function shown in equation 3.5 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. Hence, $r_{\mathrm{DES}}(z)$ and $r_{\mathrm{REQ}}(z)$ can be expressed as

$$\begin{aligned} r_{\mathrm{DES}}(z) &= r_{\mathrm{DES_{REF}}}(z) - r_{\mathrm{DES_{ERR}}}(z), & (3.7)\\ r_{\mathrm{REQ}}(z) &= r_{\mathrm{REQ_{REF}}}(z) - r_{\mathrm{REQ_{ERR}}}(z). & (3.8)\end{aligned}$$

The actual input and output rates are found out from the time-delay models as

$$\begin{aligned} r_{\mathrm{IN}}(z) = r_{\mathrm{DES}}(z) z^{-\alpha} - d_{\mathrm{BW}}(z), & \quad (3.9)\\ r_{\mathrm{OUT}}(z) = r_{\mathrm{REQ}}(z) z^{-\beta} - d_{\mathrm{FW}}(z). & \quad (3.10)\end{aligned}$$

The delays in equations 3.9 and 3.10 can be modelled 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, limiters can be used. From equation 2.2 and input and output rates shown above, the instantaneous queue length, $q(z)$, will thus grow as

$$q(z) = \frac{T_s}{z-1}\left[r_{\mathrm{IN}}(z) - r_{\mathrm{OUT}}(z)\right]$$
$$= T_s \underbrace{\frac{z^{-1}}{(1-z^{-1})}}_{\text{IIR filter}} \left[r_{\mathrm{IN}}(z) - r_{\mathrm{OUT}}(z)\right]. \quad (3.11)$$

The IIR filter implementation (functioning as an integrator) in equation 3.11 needs only two taps and is stable, since only one pole exists on the unit circle. 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 β samples. An integrated filter-based solution to the proposed congestion controller is shown in Fig. 3.1.

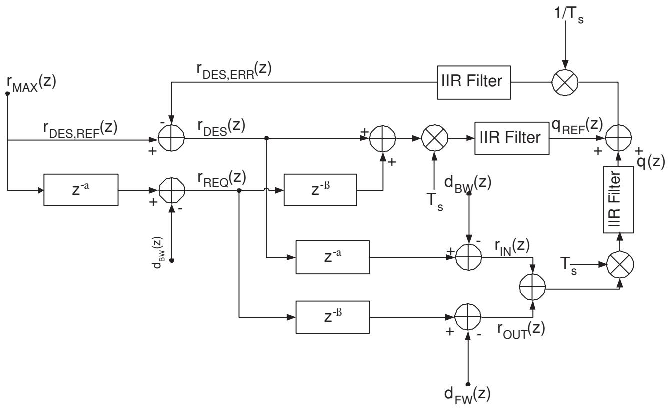

Fig. 3.1. Integrated filter-based congestion controller

4 Simulations and Results

Assuming fluid flow approximation for data in the WLAN for the controller, simulations are carried out to test for controllability of the system described in Sec. 3. 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. 2 and in Sec. 3, can be directly placed in the bridge controller. 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 $r_{\mathrm{REQ}}(z)$.

4.1 Simulation in WLAN

Based on the derivations in Sec. 3, the following parameter values are assumed for simulation: $T_s = 0.001s$, $r_{\mathrm{MAX}}(z) = 11Mbps$, $T_{\mathrm{BW}} = 100T_s$, $T_{\mathrm{FW}} = 200T_s$, $T_{\mathrm{K}} = 100T_s$. The choice of T_s is such that forward and backward delays can be effectively implemented using delay filters, as mentioned before. 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 300 each for the delays which can be implemented by filters shown in Sec. 3. In all the following simulations, the initial queue length is assumed to be of zero length.

Using the scenario in the Fig. 2.1, the simulation is performed using a MATLAB Simulink model of the filter based controller. 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 the same maximum rate, $r_{\mathrm{MAX}}(z) = 11Mbps$. The start and stop time for the connections are shown below in Tab. 1

Connection	Start Time, s	Stop Time, s
$A \rightarrow M$	0.2	5.0
$C \rightarrow L$	2.0	5.5
$G \rightarrow K$	1.8	10
$F \rightarrow J$	5.0	9.5

Table 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. 1, no transmission takes place upto $0.2s$ and as such the queue lengths increase linearly as $11Mbps \times 0.2s = 2.2Mbits$ according to equation 3.3. 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 using frequency hopping, and node H suffers disturbances of $d_{\mathrm{BW}}(t) = 8.8Mbps$ and $d_{\mathrm{FW}}(t) = 0Mbps$ as shown in Fig. 4.1. During this interval, the controller copes to mitigate the disturbance through decreasing $r_{\mathrm{DES}}(t)$ and increasing $r_{\mathrm{REQ}}(t)$ as shown in Fig. 4.2. This causes $q(t)$ to fall down with the reference queue length due to effect described in equation 3.11 and 3.2. Eventually $r_{\mathrm{IN}}(t)$ also falls down to $11Mbps \times 0.2s = 2.2Mbps$ compared to $r_{\mathrm{OUT}}(t)$ during $T_{\mathrm{FW}} = 0.2s$ interval as shown in Fig. 4.3.

At $t = (2s + T_{\mathrm{BW}})$, $d_{\mathrm{BW}}(t)$ falls down to $0.12Mbps$, while $d_{\mathrm{FW}}(t) = 9.2Mbps$. This time the effect of forward disturbance becomes predominant. To mitigate the effect of $d_{\mathrm{FW}}(t)$, the compensation through $r_{\mathrm{DES_{ERR}}}(t)$ increases and $r_{\mathrm{DES}}(t)$ drops down according to equation 3.7. Because no compensation applied through $r_{\mathrm{REQ}}(t)$, $r_{\mathrm{IN}}(t)$ falls to $2.2Mbps$ exponentially according to equation 3.5. This causes $q(t)$ to exponentially grow upto $4.0Mbits$ according to equation 3.11. After maximum compensation is reached through $r_{\mathrm{DES_{ERR}}}(t) = 7.0Mbps$, the queue length remains constant. The forward disturbance is constant until connection $A \rightarrow M$ drops with transmission completed between them at $t = 5s$. Due to change of $q_{\mathrm{REF}}(t)$ according to equation 3.3, the error filter tries to compensate the error through reducing $r_{\mathrm{DES}_{\mathrm{ERR}}}(t)$ from $7.0Mbps$ exponentially. After $t = (5 + T_{\mathrm{BW}} + T_{\mathrm{FW}})$, the compensation is complete and $r_{\mathrm{DES}}(t)$ and $r_{\mathrm{REQ}}(t)$ are the same. However, the connection $F \to J$ starts to transmit at this time. The connections $C \to L$, $G \to K$ and $F \to J$ transmit through H at the same time with no disturbance being generated until $t = 5.5s$. For the next transmissions again, no disturbance is generated, as well and as such, $q(t)$ follows $q_{\mathrm{REF}}(t) = 2.2Mbits$, while also $r_{\mathrm{IN}}(t) = r_{\mathrm{OUT}}(t) = r_{\mathrm{MAX}}(t)$. This happens in accordance to control objective 2. According to equation 2.14, the maximum buffer length is $S = 11Mbps \times (T_{\mathrm{FW}} + T_{\mathrm{BW}} + T_K) = 11Mbps \times (0.2 + 0.1 + 0.1) = 4.4Mbits$. This is the maximum queue length when $d_{\mathrm{FW}}(t) = r_{\mathrm{MAX}}(t) = 11Mbps$. In this simulation the maximum queue length reaches $4Mbits$, when $d_{\mathrm{FW}}(t) = 9.2Mbps$. System stability has already been shown during the derivations in Sec. 3. The proposed model can be used real-time system as a hardward solution, rather than as a demand-basis system. The rates are updated every T_s seconds without causing a problem to the system scalability as long as the rate updating is done by the system at times, which are integer multiple of T_s. Owing to fast hardware available to date, the proposed scheme can also lead to extremely fast solutions. Every connection at the bridge would need a separate filter-based controller, while source and destination nodes would need only one. Based on a fairness criterion, the maximum number of connections through a bridge and hence filters can be limited to minimize the cost.

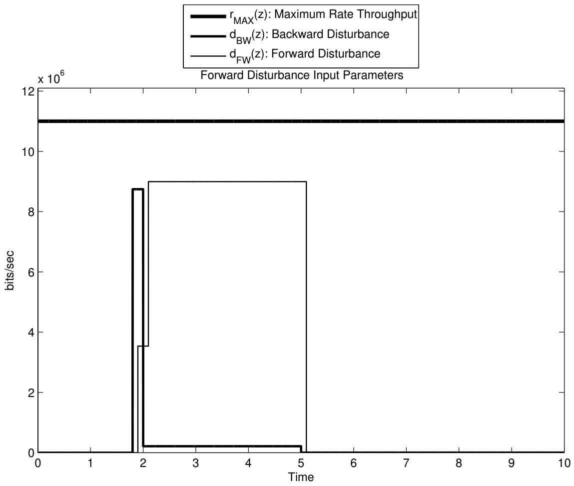

Fig. 4.1. Maximum throughput and disturbances

4.2 Comparisons

In [9], 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 [9], this has not been done. This leads to

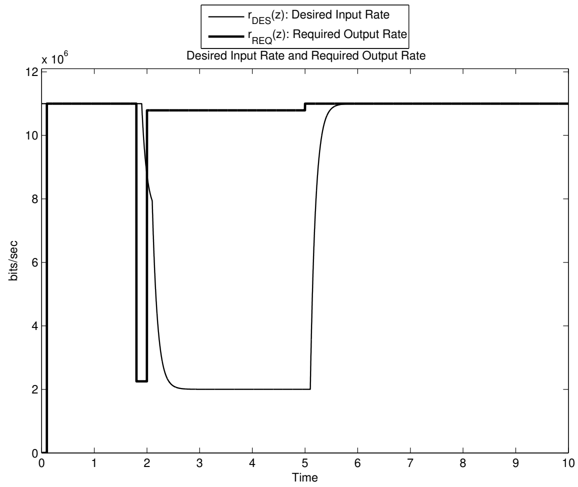

Fig. 4.2. Desired and required rates

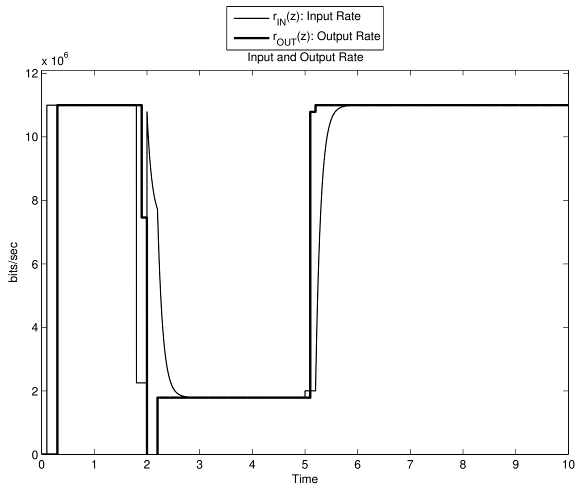

Fig. 4.3. Input and output rates

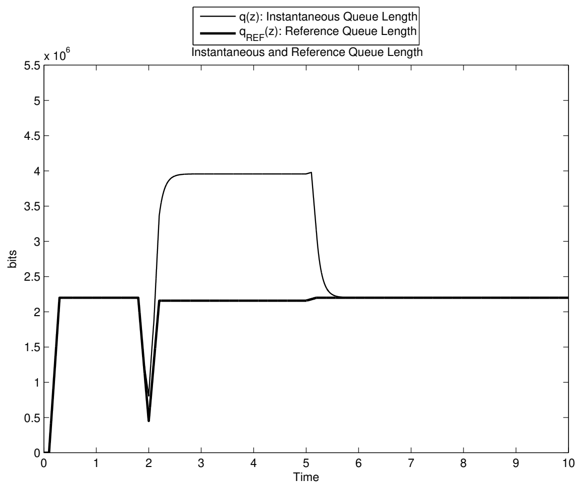

Fig. 4.4. Reference and instantaneous queue length

impractical control when either $r_{\mathrm{IN}}(t)$ or $r_{\mathrm{OUT}}(t)$ falls below zero, due to compensation of $r_{\mathrm{DES}}(t)$ through $r_{\mathrm{DES_{ERR}}}(t)$.

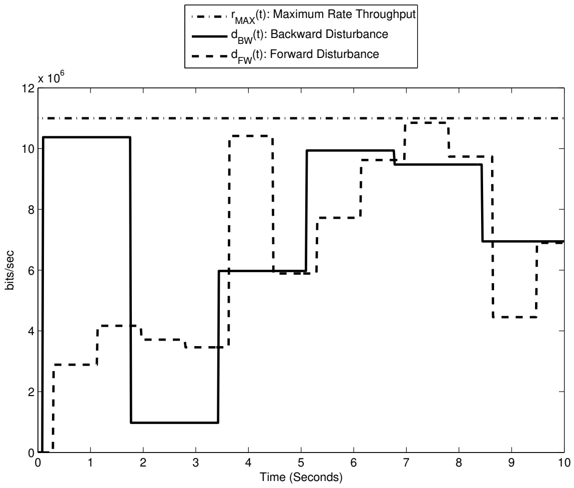

Fig. 4.5. Disturbances and maximum throughput

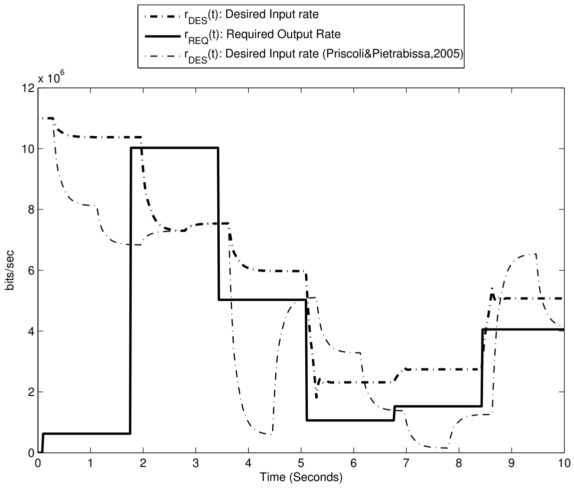

Fig. 4.6. Desired and required rates

To illustrate the differences, simulations are carried out with the proposed model and the model described in [9], whereby both are shown in the time-domain. Fig. 4.5 shows the disturbances considered over a time duration of $10s$. With the introduction of the logic for the lower limit of $r_{\mathrm{IN}}(t)$ and $r_{\mathrm{OUT}}(t)$ as zero, the control objectives are still met and remains practical as shown in the Fig. 4.6 to Fig. 4.8. Note that at $t = 3.5s$, $q(t)$ does not rise exponentially as the controller externally sets the lower value as $r_{\mathrm{IN_{MIN}}}(t) = 0$ instead of negative value, while according to [9], the rates go negative, which is not realistic. The effect of negative rates on the queue lengths are shown in Fig.s 4.7 and 4.8 along with the realistic assumptions.

In our proposed method, we also consider a realistic maximum buffer length compared to [10], where the buffers are considered unrealistically large to contain all the receive traffic. The minimum buffer size that allows the proposed control system to

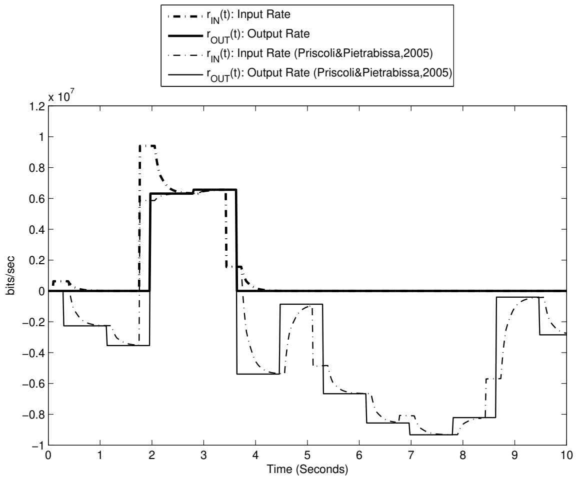

Fig. 4.7. Input and output rates

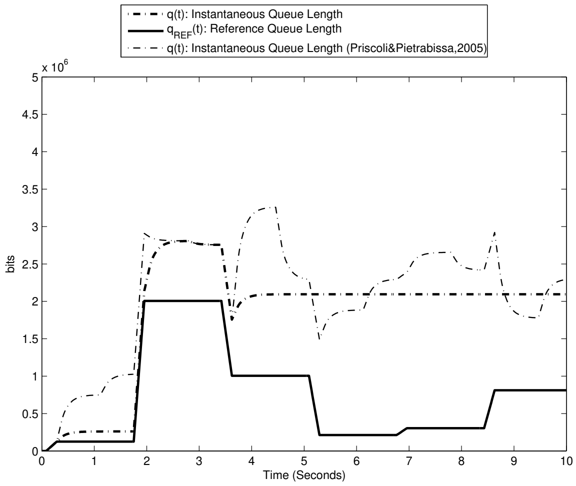

Fig. 4.8. Reference and instantaneous queue length

work confroming with the control objectives has been defined in equation 2.18.

5 Conclusions

The proposed model 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 groundbreaking 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.

References

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

[3] 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.

[4] 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.

[5] A. Goldsmith. *Wireless Communications*. Cambridge University Press, Stanford University, 1st edition edition, 2005.

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

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

[8] 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.

[9] F. D. Priscoli and A. Pietrabissa. "Hop-by-Hop Congestion Control for Wireless LAN: A Model-Based Control Approach". *International Journal of Control*, 78(12):887–905, 15 August 2005.

[10] 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.

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

[12] 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.