Providing Integrated QoS Control for IEEE 802.16 Broadband Wireless Access Systems

Jianfeng Chen, Wenhua Jiao, Qian Guo
Lucent Technologies, Bell Labs Research China
{chenjf, wjiao}@lucent.com

Abstract- **This paper proposes a new integrated QoS architecture for IEEE 802.16 Broadband Wireless MAN in TDD mode. The current strategy of providing IntServ via IEEE802.16-2004 WirelessMAN is analyzed. A mapping rule for providing DiffServ between IP layer and MAC layer is given. A fast signaling mechanism is designed to provide cross layer integrated QoS for Point to Multi-Point (PMP) mode. Comparison between traditional RSVP signaling mechanism and proposed signaling mechanism is given. The simulation is conducted for VoIP, FTP and HTTP traffic. The results show that the proposed integrated QoS control mechanism can satisfy the bandwidth requirements of different applications.**

Key Words: IEEE 802.16, broadband wireless access, Integrated QoS

I. Introduction

QoS refers to traffic control and resource management mechanisms that guarantee performance to applications, traffic flows, or packets. Since IP network service is based on a connectionless and best-effort model, this service model is not adequate for many applications that normally require assurances on QoS performance metrics such as packet latency, jitter, packet loss rate and bandwidth. So a number of enhancements have been proposed to enable offering different levels of QoS in IP networks including the integrated services (IntServ) architecture [1], the differentiated service (DiffServ) architecture [2].

IntServ is implemented by four components: the signaling protocol (e.g. RSVP), the admission control, the classifier and the packet scheduler. Applications requiring guaranteed service or controlled-load service must set up the paths and reserve resources before transmitting their data. The admission control routines will decide whether a request for resources can be granted. After classification of packets in a specific queue, the packet scheduler will then schedule the packet to meet its QoS requirement.

RSVP is a signaling protocol for providing QoS in the Internet. The sender sends a PATH message to the receiver to specify the characteristics of the traffic. The receiver responds with a RESV message to reserve resources for the flow. The flow should be periodically refreshed. The service flow can be either automatically removed along with refresh timeout or explicit deleted using PATHTEAR or RESVTEAR message from source or destination respectively.

IEEE 802.16 WirelessMAN can support multiple types of communication services (data, voice, video) with different QoS requirements. QoS signaling mechanisms and functions are defined in MAC layer to control data transmission between BS and SS. As defined in [3], there are 4 types of MAC layer services characterized by QoS parameters such as latency, jitter, throughput, minimum reserved traffic rate, maximum sustained traffic rate. These service flows can be created, changed, or deleted through the issue of Dynamic Service Addition (DSA), Dynamic Service Change (DSC), and Dynamic Service Deletion (DSD) messages. Each of these actions can be initiated by the SS or the BS and are carried out through a two or three-way-handshake. For example, a new service flow initiated by the SS is built as follows. When SS detects the emergence of a new service flow, it will calculate the available resources to determine whether a DSA request will be sent or not. Upon reception of the DSA request, the BS verifies whether this request can be supported, and sends a DSA response. Finally, according to the message of the DSA response, the SS sends a DSA acknowledgement and enables the new service flow if the request is approved. In [3], some rules to classify DiffServ IP packets into different priority queues are also proposed based on IP QoS indication bits in IP header. So, in general, the QoS architecture of IEEE 802.16-2004 under PMP mode can support both IntServ and DiffServ.

In previous work [4-6], different packet scheduling algorithms with QoS support are proposed for IEEE802.16 WirelessMAN. They all concentrate on the QoS problem in MAC Layer. However, how to guarantee the high layer QoS for services traversing WirelessMAN MAC and PHY layer hasn't been addressed.

Two ways for providing cross layer QoS control via WirelessMAN technology may be candidates:

For the first one, the traditional RSVP is used to provide cross layer QoS control. RSVP signaling message will be regarded as the traffic data by convergence sub-layer in the MAC layer.RSVP signaling message can be classified into a special high priority queue (refer to protocol queue below) by a protocol-specific packet matching criteria. This RSVP queue will be transmitted in the second management connection. Other protocol-specific packet such as Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP), SNMP, etc. also transmitted using this connection. In summary, the QoS provision procedure will consist the following two part: on one hand, the secondary management connection will be used for RSVP to provide the layer 3 QoS; on the other hand, the primary management connection will be used for DSA/DSC/DSD to provide the layer 2 QoS. Since the second management connection is defined for delay tolerant traffic and there are many other IP protocol related message (DHCP, SNMP, TFTP, etc) sharing the same queue, the whole QoS provision will be rather slow. Furthermore, considering the RSVP signaling has a periodical refreshing procedures, which consume a lot of bandwidth, it is not efficient to use the secondary management connection for the RSVP.

The second one is the proposed way. Since there are so many similarities between providing Internet IntServ using RSVP and MAC layer QoS using DSA/DSC/DSD, naturally, the mechanism will be superior to the first one in high efficiency and fastness by mapping between the cross layer QoS control and MAC QoS in IEEE802.16 network. The paper is organized as follows. In Section 2 we provide the architecture for integrated QoS control in details. In section 3, we give a comparison between the two ways of RSVP. Section 4 provides simulation result of our proposed architecture. Section 5 concludes the paper.

II. Architecture for integrated QoS control

In order to provide multi-layer QoS control, a convergence sub-layer is defined in IEEE 802.16 protocol [3] to accept higher-layer protocol data units and perform classification and mapping function. Although some parameters such as source/destination IP address, source/destination port and protocol are defined to fulfill IP packets' classification, it is impossible to acquire necessary value of bandwidth and delay for dynamic service request or bandwidth request. Another deficiency in this protocol is lack of support for IntServ services.

Since the communication under PMP mode in IEEE 802.16 is connection-oriented, the application must establish the connection with the BS as well as the associated service flow (UGS, rtPS, nrtPS or BE). BS will assign the connection with a unique connection ID (CID) to each uplink or downlink transmission. The message exchange for DSA and DSC can be deployed to achieves QoS guarantees through end-to-end (SS-to-BS and BS-to-SS) resource (bandwidth/buffer) reservation for packet flows and performs per-flow scheduling which IntServ services require. For DiffServ services, on the other hand, a number of per-hop behaviors (PHBs) for different classes of aggregated traffic can be mapped into different connections directly. The requirements of a multi-layer integrated QoS control architecture may include: (1) Guarantee different level QoS; (2) Prioritize the traffic classes; (3) Conduct multi-granularity traffic grooming efficiently; (4) Adjust resource allocation dynamically; (5) Share resources fairly. To meet all these requirements, we propose an integrated QoS control architecture as shown in Fig. 1, which implements a cross layer traffic-based prioritization mechanism in a comprehensive way.

Step 1 and 2 in Fig. 1 show when a new service flow arrives in IP layer, it will be firstly parsed according to the definition in PATH message (for InteServ) or Differentiated Services Code Point (DSCP for DiffServ); then classified and mapped into one of four types of services (UGS, rtPS, nrtPS or BE). The detailed explanation for traffic classification and mapping strategies is given in the following section. In step 3, the dynamic service model in SS will send request message to the BS, then the admission control in BS will determine whether this request will be approved or not. If not, the service module will inform upper layer to deny this service in step 4; if yes, admission control will notify scheduling module to make a provision in its basis scheduling parameter according to the value shown in the request message and the accepted service will transfer into traffic grooming module in step 5. According to the traffic grooming result, SS will send Bandwidth Request message to BS in step 6. The scheduling module in BS will retrieve the requests (step 7) and generate UL-MAP and DL-MAP message (step 8) following the bandwidth allocation results. Finally, the SS will package SDUs from IP layer into PDUs and upload them in its uplink slot to SS (step 9-10).

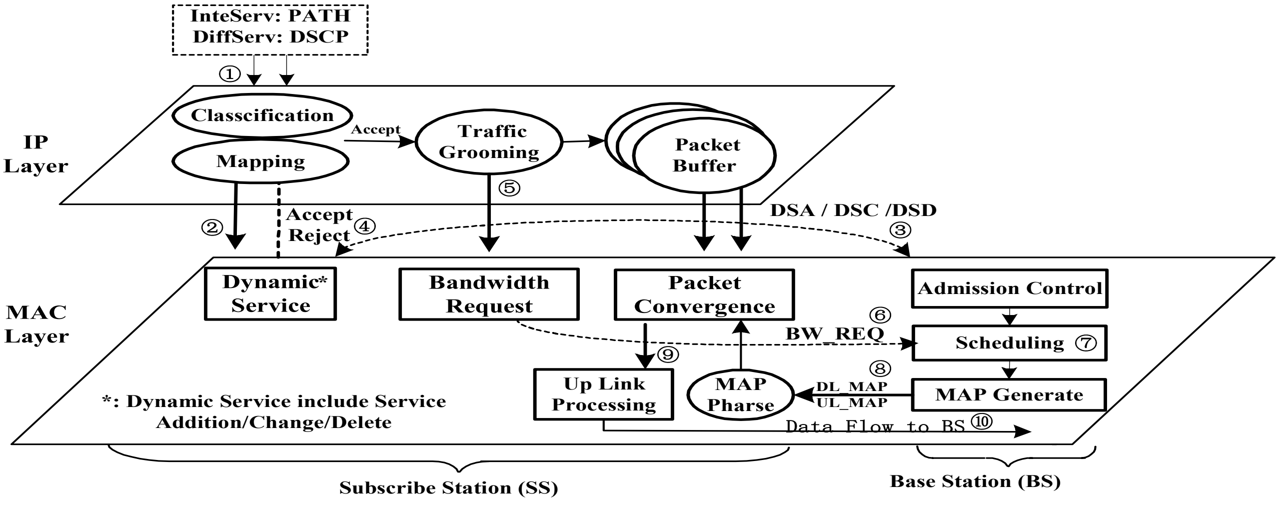

Fig.1. The multi-layer integrated QoS control architecture

Table 1. Mapping Rules for IntServ Services

Traffic Class	Bandwidth Requirements	Delay / Jitter / Loss Rate	MAC layer Services
Hard QoS guarantee(eg. VPN tunnel, Leased line E1/T1)	*Constant bandwidth*	*Minimum packet delay, jitter and loss rate*	Unsolicited Grant Service
Soft QoS guarantee(eg. VoIP, VOD, digital TV, FTP, gaming.)	*Guaranteed*	*Regular delay, jitter require*	Real-Time Polling Service
	Not guaranteed	*long delay, jitter require*	Non-Real-Time Polling Service
Best effort (eg. HTTP)	*Only basic connection*	*N/A*	Best Effort

Table 2. Mapping Rules for DiffServ Services

Traffic Class	Service Description	DS Octet (DS5-3)	MAC layer Services
Hard QoS guarantee(eg. VPN tunnel, Leased line E1/T1)	*Critical*	*101*	Unsolicited Grant Service
Soft QoS guarantee(eg. VoIP, VOD, digital TV, FTP, gaming.)	*Flash Override, Flash, Immediate*	*100 / 011/010*	Real-Time Polling Service
	Priority,	*001*	Non-Real-Time Polling Service
Best effort (eg. HTTP)	*Runtime*	*000*	Best Effort

A. Traffic Classification and Mapping Strategies for IntServ Services

As shown in Fig. 2, the sender will send a PATH message including traffic specification (TSpec) information. The parameters such as up/bottom bound of bandwidth, delay and jitter can be easily mapped into parameters in DSA message such as Maximum Sustained Traffic Rate, Minimum Reserved Traffic Rate, Tolerated Jitter and Maximum Latency. According to the response of DSA message, the provisioned bandwidth can be also freely mapped into reserved specification (RSpec) into RESV message. As illustrated in Table 1, four rules are defined to map IP layer service into MAC layer services.

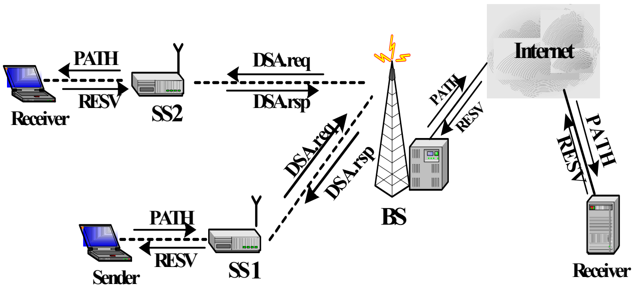

Fig. 2. Traffic Classification and Mapping for IntServ Services

B. Traffic Classification and Mapping Strategies for DiffServ Services

For DiffServ services, DSCP code is deployed for classification. As shown in Fig. 3, the first 3 bits are for class selector, the middle 3 bits are for drop priority. There are three definitions of per-hop behavior (PHB) to specify the forwarding treatment for the packet.

ToS Octet	P2	P1	P0	T3	T2	T1	T0	Zero
DS Octet	DS5	DS4	DS3	DS2	DS1	DS0	ECN1	ECN0
	(Class Selector)			(Drop Precedence)				

Fig. 3. Differentiated Services Code Point

Expedited forwarding (EF) [7] is intended to provide a building block for low delay, low jitter and low loss services by ensuring that the EF aggregate is served at a certain configured rate.

Assured Forwarding (AF) [8] PHB group is to provider different levels of forwarding assurances for IP packets. Four AF classes are defined, where each AF class is allocated a certain amount of forwarding resources (buffer space and bandwidth). As illustrated in Table 2, four rules are defined to map IP layer service into MAC layer services.

C. Admission Control and Scheduling in BS

It will collect all the DSA/DSC/DSD requests and update the estimated available bandwidth (C_a) based on bandwidth change. Suppose there are I classes of service and the i^{th} classes of service has totally J_i connection, the available bandwidth equals to:

$$C_a = C_{tatal} - \sum_{i=0}^{I} \sum_{j=0}^{J_i - 1} r_{\min}(i,j) \qquad (1)$$

In which $r_{min}(i,j)$ is the Minimum Reserved traffic rate of the j^{th} connection in the i^{th} class of service flow. For those connections whose r_{min} equal to zero, they can always be accepted by our admission policy. But the QoS of these connections will not be guaranteed anytime. They always have the lowest priority. Their connections will be interrupted anytime unless the QoS requirements of all other connections can meet sufficiently.

When a new service flow comes or an old service flow requests to change its QoS, the following principle should hold:

$$C_a \geq 0 \qquad (2)$$

In this paper, eq. (2) is the QoS policy for admission control.

The hierarchical structure of the bandwidth allocation in BS is shown in Fig. 4. In this architecture, two-layer scheduling is deployed. Six queues are defined according to their direction (uplink or downlink) and service classes (rtPS, nrtPS and BE). Since service of UGS will be allocated fixed bandwidth (or fixed time duration) in transmission, we will cut these bandwidths directly before each scheduling.

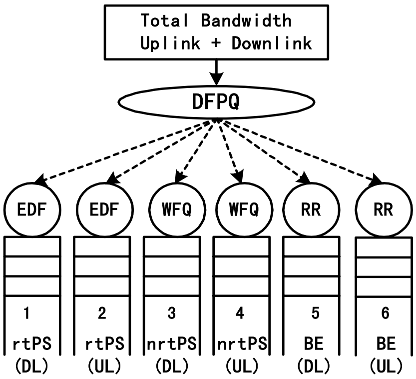

Fig. 4 Hierarchical structure of bandwidth allocation

The algorithm of the first layer scheduling is called Deficit Fair Priority Queue (DFPQ) proposed in [4], which is basically based on priority queue. 2 policies of initial priority are defined as following:
Service class based priority: rtPS > nrtPS> BE
Transmission direction based priority: Downlink > Uplink.

In the second layer scheduling, three different algorithms are assigned to three classes of service to match its requirements. We apply earliest deadline first (EDF) for rtPS [9], which means packets with earliest deadline will be scheduled first. The information module determines the packets' deadline and the deadline is calculated by its arrival time and maximum latency.

Weight fair queue (WFQ) [10] is deployed for nrtPS services. We schedule this type of packets based on the weight (ratio between a connection's nrtPS Minimum Reserved Traffic Rate and the total sum of the Minimum Reserved Traffic Rate of all nrtPS connections). The remaining bandwidth is allocated to each BE connection by round robin (RR).

III. Comparison between two ways of RSVP

In this section, two ways of providing cross layer QoS control via WirelessMAN technology are compared. For brevity, we only describe the process between BS and SS1 in Fig. 2, and it is preventative. When a service flow is needed from Sender to Receiver, the processes of the RSVP between BS and SS1 are illustrated in Table 3 and Table 4 respectively.

As shown in Table3, the negotiation of QoS parameters for one traffic will be processed two times. For the first time, the parameters are carried in RSVP messages and transmitted through the Secondary Management connection. For the second time, the same parameters are mapped in MAC message and transmitted through the Primary Management Connection. it is not a good way obviously.

However, Table 4 shows the RSVP signaling messages are mapped directly into the MAC messages, and then transmitted through the Primary Management Connection. In this way, the messages are transmitted only once ingeniously, reducing the delay.

From the analysis, by avoiding the redundancy transmission of the same RSVP message the proposed way of RSVP is superior to traditional way in high efficiency and fastness for providing cross layer QoS control via IEEE802.16 MAC layer QoS mechanism.

Table 3 Traditional way of RSVP

SS		**BS**
Service Flow “Sender -->Receiver”		
Received PATH Check if resource are available Update the state Send PATH	PATH →	Received PATH Check whether IP QoS are available Update the state Transfer PATH to next hop
		Received RESV Reserve Resource of IP QoS
Received RESV Reserve Resource Map IP QoS to MAC QoS	←RESV	Transfer RESV
Send DSA-REQ	DSA-REQ →	Received DSA-REQ
Received DSA-RSP Transfer RESV to previous hop	←DSA-RSP	MAC Admission Control Send DSA-RSP

Table 4 Proposed way of RSVP

SS		**BS**
Service Flow “Sender →Receiver ” Received PATH Check if resource are available MAC layer QoS mechanism		
Map PATH to MAC QoS Send DSA-REQ	DSA-REQ →	Received DSA-REQ MAC Admission Control Send DSA-REQ message to next hop
Received DSA-RSP Map DSA-RSP to RESV RESV received by IP Layer Transfer RESV to previous hop	DSA-RSP ←	Received DSA-REQ from the next hop MAC Admission Control Send DSA-RSP

IV. Simulation Results

We have developed simulation platform for the proposed integrated QoS architecture. The simulation environment consists of four-node topology including one BS and three SSs, operating in IEEE 802.16 PMP mode. Suppose that, between BS and each SS, there will be one or more service flow for each kind of Service in uplink or downlink. We follow the mapping rules defined in Table 1 and 2, the VoIP service is mapped into rtPS service; the FTP service is mapped into nrtPS service; and the HTTP service is mapped into BE service. The total bandwidth and the duration for each frame are assumed to be 10Mbit and 10ms respectively, so the bandwidth for a frame is 100Kbit. In the simulation all packet arrivals occur at the beginning of each frame and the packet arrival process for each connection follows the Poisson distribution with different traffic rate λ. Each connection has specific QoS parameters in terms of Maximum Sustained Traffic Rate, Minimum Reserved Traffic Rate and Maximum Latency requirement.

UGS flow request 5Mbit/s in uplink and 5Mbit/s in downlink with hard QoS with constant bandwith. In the simulation, UGS flow is allocated constant bandwith in each frame. So there will be only 90Mbit/s bandwith left. The other types of service flows used in the simulation are given in Table 5.

Table 5. Input Service Flow

Service Type	Mapped Type	CID	Average Bandwidth (Kbit)	Max. Delay (ms)	Max.sustained traffic rate (Kbit)	Min.reserved rate (Kbit)
VoIP	DL_rtPS	1	10	60	12	8
		2	10	40	12	8
		3	10	20	12	8
VoIP	UL_rtPS	4	7	70	8.4	5.6
		5	7	50	8.4	5.6
		6	6	30	7.2	4.8
FTP	DL_nrtPS	7	5	100	6	4
		8	5	100	5	4
		9	5	100	5	4
FTP	UL_nrtPS	10	4	100	6	4
		11	4	100	5	4
		12	4	100	5	4
HTTP	DL_BE	13	2	240	-	1.6
		14	2	240	-	1.6
		15	2	240	-	1.6
HTTP	UL_BE	16	2	300	-	1.6
		17	1	300	-	0.8
		18	1	300	-	0.8

As shown in Fig. 5, the bandwidth allocated for each type of service fits the traffic rate of data source. Each service curve adapts and follows the data source curve.

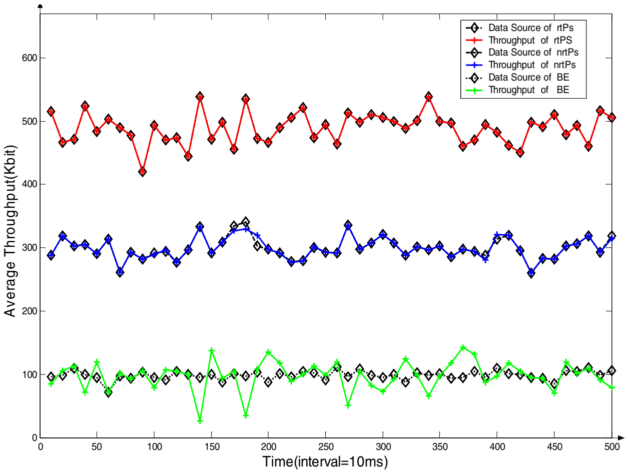

Fig. 5. Input data source vs. traffic throughput

We also examine the effect of throughput of other existing services when one type of new service flows join the network. The scenario is the number of nrtPS services rising from 0 to 8(bandwith rising from 0 to 40Kbit/frame), resulting in the total bandwidth rising from 70Kbit/frame to 110Kbit/frame), while other service flows keep unchanged. As shown in Fig. 6, when the total traffic rate exceeds the total bandwidth in a frame, the throughput of rtPS is unchanged, but the throughput of BE decreases to its minimum reserved rate. The result shows when new service flows enter, the proposed integrated QoS architecture guarantees the existing bandwidth for the high priority services as well as the minimum reserved rate for low priority services.

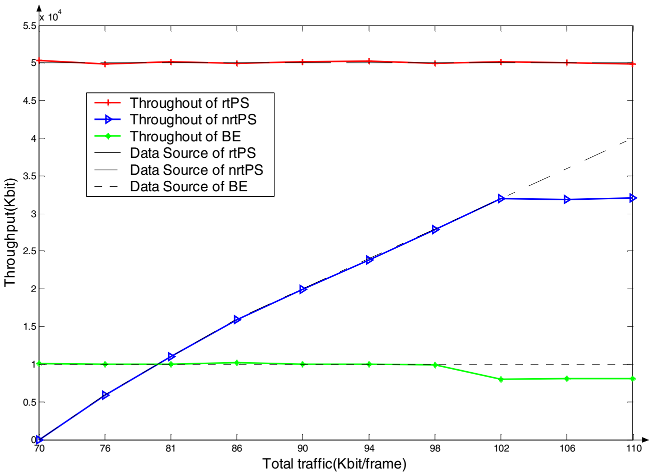

Fig. 6. Throughput when new flows (nrtPS) joins

V. Conclusion

In this paper, we propose an architecture to provide multi-layer QoS control for IEEE 802.16 WirelessMAN in PMP mode. Both IntServ and DiffServ are supported. Compared with the traditional way of providing cross layer QoS via WirelessMAN, the proposed integrated QoS control is superior in high efficiency and fastness to guarantee the throughput requirements of source traffic.

Reference

[1] R. Braden, D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: an Overview,” IETF RFC 1633, June 1994.

[2] S. Blake *et al.*, “An Architecture for Differentiated Services,” IETF RFC 2475, Dec. 1998.

[3] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, October 2004

[4] Jianfeng Chen, Wenhua Jiao, Hongxi Wang, “A Fair Scheduling for IEEE 802.16 Broadband Wireless Access Systems”, ICC2005, May 16-20, Souel, Kerea

5] Hawa, M.; Petr, D.W.,Quality of service scheduling in cable and broadband wireless access systems, Tenth IEEE International Workshop on Quality of Service, p247-255, 2002

[6] K. Wongthavarawat, and A. Ganz, “Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems”, International Journal of Communication Systems, Vol. 16, p81-96, 2003

[7] B. Davie *et al.*, “An Expedited Forwarding PHB (Per-Hop Behavior),” IETF RFC 3246, Mar. 2002.

[8] J. Heinanen *et al.*, “Assured Forwarding PHB Group” IETF RFC 2597, June. 1999.

[9] Georgiadis L, Guerin R, Parekh A. Optimal Multiplexing on a Single Link: Delay and Buffer Requirements. Proceedings of IEEE INFOCOM 94; vol. 2, 1994; 524–532.

[10] Demers A, Keshav S, Shenker S. Analysis and Simulation of a Fair Queuing Algorithm. SIGCOMM CCR 19 1989;