

Quality of Service Control for WLAN-based Converged Personal Network Service

E.-C. Park, I.-H. Kim, G.-M. Jeong, B. Moon

Eun-Chan Park, Bongkyo Moon

Department of Information and Communication Engineering,
Dongguk University-Seoul, Korea
E-mail: ecpark@dongguk.edu, bkmoon@dongguk.edu

In-Hwan Kim

Mobile Device Development Team,
SK Telecom, Seoul, Korea
E-mail: ihkim@sktelecom.com

Gu-Min Jeong

School of Electrical Engineering,
Kookmin University, Seoul, Korea
E-mail: gm1004@kookmin.ac.kr

Abstract: This paper proposes a framework for quality of service (QoS) control in WLAN-based converged personal network service (CPNS). First, we show that the CPNS devices in WLANs occupy the shared wireless channel in an unfair manner; and thus, QoS is degraded. The reasons of such problem are analyzed from two viewpoints of (i) channel access mechanism according to carrier sensing multiple access protocol of WLAN and (ii) TCP congestion control mechanism in response to packet loss. To improve QoS and assure fair channel sharing, we propose an integrated QoS control framework consisting of admission control and rate control. Based on the available capacity, the admission control determines whether or not a specific QoS service can be admitted. The rate control is implemented using congestion window control or token bucket algorithm. The proposed mechanism differentiates QoS service from best-effort (BE) service such that the QoS service is preferentially served to satisfy its QoS requirements and the BE service is served to share the remaining resource in a fair manner. The extensive simulation results confirm that the proposed scheme assures QoS and fair channel sharing for WLAN-based CPNS.

Keywords: CPNS, QoS, fairness, WLAN, TCP

1 Introduction

Recently, many new portable devices equipped with Wi-Fi interface, e.g., smart phone, tablet PC, mobile internet device (MID), portable media player, and digital camera/camcorder, have been introduced. They are connected to the Internet or with each other and enable many new emerging services to come into wide use, such as real-time multimedia streaming, video telephony, on-line game, image/video sharing. In order to provide converged service for the portable devices in short-range personal networks such as home networks or in-car networks, Open Mobile Alliance (OMA) makes a standard for converged personal network service (CPNS) [1], [2] by identifying the overall service architecture, service scenario, and functional requirements. The CPNS system consists of CPNS server, personal network gateway (PN GW), and personal network element (PNE); the PNE accesses to the PN GW via wireless link and is provided with

various services from the CPNS server. We consider in this paper that the CPNS system is implemented with IEEE 802.11 wireless local area networks (WLANs) [3]. WLAN is widely deployed in home networks and installed in many devices, i.e., WLAN is a *de facto* standard for wireless networking for portable devices in small area, and it is the most suitable wireless access technology considering the service scenario, coverage, and data rate for CPNS.

This paper deals with the issue of QoS assurance in WLAN-based CPNS. The various services of CPNS (e.g., real-time multimedia streaming, Internet telephony, file uploading/downloading, remote surveillance) have different QoS requirements in terms of delay, loss, or throughput; however it is a challenging problem to satisfy these diverse QoS requirements because the MAC protocol of WLAN works in a distributed contention-based manner and the quality and capacity of wireless channel are time-varying and location-dependent. First, we show that the PNEs¹ access the channel in an unfair way, i.e., the uploading (UL) PNEs dominate the use of shared channel while the downloading (DL) PNEs cannot occupy their fair shares; and thus, QoS can be severely degraded especially for DL PNEs. We analyze the reasons of this problem from two viewpoints; (i) unfair channel access mechanism of MAC and (ii) asymmetric congestion control mechanism of TCP.

To resolve this problem, we propose an integrated QoS control framework, which consists of admission control and rate control. We classify the service into two categories; service with and without QoS requirements, each of which is referred to as *QoS service* and *best-effort (BE) service*, respectively. The proposed scheme differentiates the QoS service from the BE service. The admission control is applied to the QoS service in order to prevent an excessive request of QoS service from being allowed. The admission control estimates the available channel capacity and converts the required data rate of a QoS service to the equivalent channel capacity. Only if the requirement is less than the available capacity, the QoS service can be admitted and the required capacity is virtually reserved. On the other hand, the rate control is applied to both QoS service and BE service so that the required data rate of QoS service can be assured and the BE services are enforced to share the channel in a fair manner. The rate control is implemented by employing TCP congestion window control mechanism or token bucket mechanism. Furthermore, the proposed framework can be extended to minimize the delay of QoS service and to support weighted fairness for BE services. Consequently, QoS can be assured for QoS services and fairness can be attained among BE services, which is confirmed through ns-2 [4] simulations.

There have been many proposals for assuring QoS and fairness in IEEE 802.11 WLANs in the literature. The enhanced distributed channel access (EDCA) mechanism has been standardized in IEEE 802.11e [5] to improve QoS by differentiating the channel access delay between different access categories. Similar QoS enhancement mechanisms have also been proposed for prioritized service differentiation [6–8]; they modify the basic MAC protocol and use differentiated MAC parameters, e.g., inter-frame space, contention window, or transmission opportunity, depending on the service priority. Moreover, the fairness issue of WLAN has been actively studied. Many MAC protocols have been proposed to assure MAC-layer fair access opportunity between competing stations or between DL and UL stations by employing a proper backoff mechanism or by controlling contention window size [9–12]. Recently, the issue of TCP fairness in WLAN has been addressed in [13–15]. They aim to assure fairness between TCP flows in WLAN by limiting TCP congestion window size [13], by employing per-flow scheduling in the access point (AP) [14], or by differentiating channel access delay based on IEEE 802.11e [15]. Compared to these previous studies, this paper makes the following contributions;

- This paper proposes a *generalized* framework to assure QoS and fairness simultaneously, it is applicable regardless of packet size, transmission rate, service direction (DL or UL), and

¹In this paper, we interchangeably use the terms, PNE, device, and station.

transport-layer protocol (TCP or UDP).

- The proposed mechanism provides QoS service with the *absolute* assurance in terms of data rate. At the same time, it provides BE service with the *weighted* fairness without decreasing the overall channel utilization.
- The proposed mechanism is *practical and simple* for CPNS; it defines a service establishment signaling but does not modify the standard IEEE 802.11 MAC protocol, and the major function of the proposed scheme needs to be implemented in the PN GW while the PNEs require a minor modification.

The rest of this paper is organized as follows. In Section 2, we identify the problem of unfair channel sharing and QoS degradation arising in WLAN-based CPNS and analyze its reasons. The QoS control framework and algorithm are proposed in Section 3, and the performance is evaluated through extensive simulations in Section 4. The conclusion follows in Section 5.

2 Problem statement

In this section, the problem of QoS degradation and unfairness is shown via the preliminary simulation, and the reasons of such problem are analyzed.

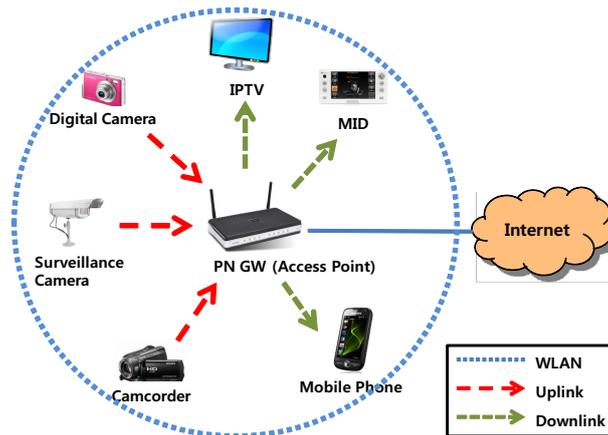


Figure 1: Service model of WLAN-based CPNS.

2.1 QoS degradation and unfairness in WLAN-based CPNS

We consider the service model of WLAN-based CPNS, especially focusing on home network, as shown in Fig. 1. The PN GW plays a role of AP in WLAN and it is connected to the Internet or CPNS server through a wired link and communicates with the PNEs. We consider the following typical service scenario for CPNS;

- the digital camera and camcorder upload their photos and moving pictures to the PN GW so that the contents can be played in the IPTV, smart TV, or digital photo frame. Also, they upload the contents to the PC or remote storage server to free up its built-in memory, or they upload the contents to social community sites (e.g., YouTube or flickr) to share them with others;
- the surveillance camera uploads the real-time streaming video to the server for the purpose of security;

Table 1: Service scenario used in the simulation.

device	service	direction	transmission protocol	required data rate	start time
IPTV	streaming	DL	TCP	1 Mb/s	0 sec
mobile phone	file transfer	DL	TCP	N/A	0 sec
camcorder	file transfer	UL	TCP	N/A	50 sec
surveillance camera	security	UL	UDP	1 Mb/s	100 sec

- the MID and mobile phone download multimedia contents from the CPNS server or the Internet to store the contents into the built-in memory and play them later. Moreover, they can play the multimedia contents in the PC or remote storage server as streaming.

We perform preliminary ns-2 simulation with the specific service scenario given in Tab. 1. Note that the service without the requirement of data rate, i.e., file transfer of mobile phone and camcorder, is considered as BE service. The MAC and PHY parameters are set according to IEEE 802.11b specification, the PHY transmission rate and packet size are set to 11 Mb/s and 1 Kbyte, respectively, the buffer size in the PN GW is set to 100 Kbyte², and the TCP window size is set to 64 Kbyte (default value in Microsoft Windows XP). The file size is assumed to be large enough for the service to last till the end of simulation. The wireless channel error is modeled as a two-state Markov-modulated process, where the packet error rate in the wireless channel is 1%.

We observe per-device throughput to evaluate the performance in terms of QoS and fairness from Fig. 2. Before the camcorder uploads a file at $t = 50$ sec, the required data rate for IPTV is satisfied and the remaining capacity is utilized by the mobile phone, and the throughputs of both devices are stable. When $t = [50 - 100]$ sec, however, the average throughput of IPTV is about 770 Kb/s, i.e., the required data rate cannot be assured, moreover its throughput fluctuates severely, which results in the significant degradation of QoS due to the increase of delay and jitter. Also, the throughput of mobile phone is decreased suddenly because of the UL file transfer of camcorder. The wireless channel capacity is not allocated in a fair way; during the time interval between 50 sec and 100 sec, the average throughput of camcorder is higher than that of IPTV and mobile phone by about 2.10 and 1.87 times, respectively. After $t = 100$ sec, the throughput of surveillance camera is stable and satisfies its required data rate of 1Mb/s and it is little affected by the other services. However, the throughputs of IPTV and mobile phone are further decreased. Even in this case, the throughput of camcorder is higher than those of IPTV and mobile phone by about 3.10 and 2.22 times, respectively. These simulation results confirm that (i) the QoS of DL service is highly vulnerable to the UL service while the QoS of UL service is little affected by the DL service, (ii) the channel capacity is unfairly shared among BE services; the UL service is favored over the DL service.

2.2 Causes of the problem

MAC: Unfair channel access

In infrastructure WLANs, all the devices are associated with the AP. The UL devices transmit packets to the AP while the DL devices receive packets from the AP. The MAC-layer contention for channel access takes place among transmitters, AP and UL devices, according to the carrier sensing multiple access (CSMA) mechanism, i.e., IEEE 802.11 distributed coordination function

²The buffer size is determined for bounded queuing delay to support the multimedia streaming services whose QoS is sensitive to delay and/or delay variation.

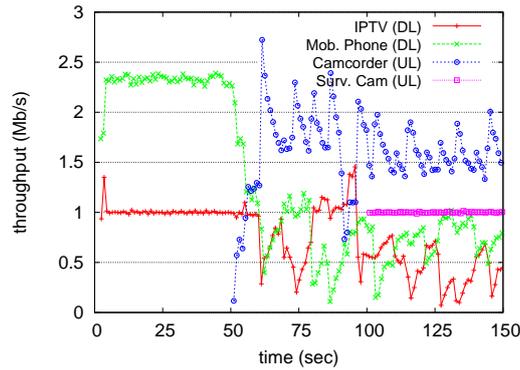


Figure 2: Per-device throughput in the WLAN-based CPNS without QoS control.

(DCF) [3]. If the AP wins the contention, the DL devices can be served, otherwise if the UL device wins, it gets its own transmission opportunity. Defining the number of UL and DL devices as N_u and N_d , respectively, the number of contending devices are $N_u + 1$, independent of N_d . The DCF mechanism provides all the contending devices with the equitable chances of channel access in the long term. Assume that all the UL devices always make attempt to send packets at the same PHY-layer transmission rate and the data generation rates for all the DL devices are same. Then, the steady-state throughputs of UL and DL devices, denoted as th_u and th_d respectively, become

$$\begin{aligned} th_u &= \frac{C}{N_u + 1}, \\ th_d &= \frac{C}{(N_u + 1)N_d}. \end{aligned} \quad (1)$$

Here, C denotes the effective capacity of WLAN. Defining a fairness index, γ , as the ratio of the average throughput achieved by the DL devices to that by the UL devices, it becomes

$$\gamma \triangleq \frac{th_d}{th_u} = \frac{1}{N_d} \quad (2)$$

from (1). It is important to note that γ depends on N_d , but it is irrespective of N_u .

We perform ns-2 simulation to validate this analysis. Figure 3(a) shows the average per-device throughput with various values of $N_d (= N_u)$ when devices send/receive UDP traffic. The discrepancy between th_u and th_d increases, i.e., γ decreases, as N_d increases. Moreover, we observe that the average per-device throughput of an individual UL device is nearly equal to the sum of per-device throughputs of all DL devices, which agrees with (1).

TCP: Asymmetric congestion control

The problem of unfair channel access becomes exacerbated when devices send/receive TCP traffic. As shown in Fig. 3(b), even when only one DL device coexists with one UL device (i.e., $N_d = N_u = 1$), the throughput of DL device is lower than that of UL device by about 26%, i.e., $\gamma = 0.74$. As N_u increases from 2 to 8, the value of γ remarkably decreases from 0.19 to 0.03.

The aggravation of unfair channel sharing results from the asymmetric congestion control mechanism of TCP in response to packet loss. In general, the capacity of WLAN is smaller than that of wired link connected to the AP. According to TCP congestion control, TCP sender increases its transmission rate until it detects a packet loss. Consequently, the downlink buffer of

AP easily becomes congested and packets are dropped due to buffer overflow. It is important to note that two different types of packets are buffered in the AP's downlink buffer; TCP data packet for DL device and TCP ACK packet for UL device. When TCP data packet for DL device is dropped in the buffer, the TCP sender recognizes this as an indication of congestion and decreases the congestion window by half to relieve the congestion; and thus the throughput of DL device is halved. On the other hand, the TCP in the UL device can be unresponsive to the TCP ACK packet loss in the AP's downlink buffer. Due to the cumulative ACK mechanism of TCP, the loss of an ACK packet for UL device does not necessarily decrease the congestion window as long as the next ACK packet with a higher sequence number is delivered before the retransmission timer expires. Therefore, the cumulative ACK mechanism lets the UL device tolerate the loss of ACK packets, and the TCP in the UL device increases the congestion window gradually until it reaches the maximum size, increasing the throughput of UL device. Consequently, the throughput of UL device becomes higher than that of DL device, and the unfairness problem occurs more severely, compared to the case of UDP traffic.

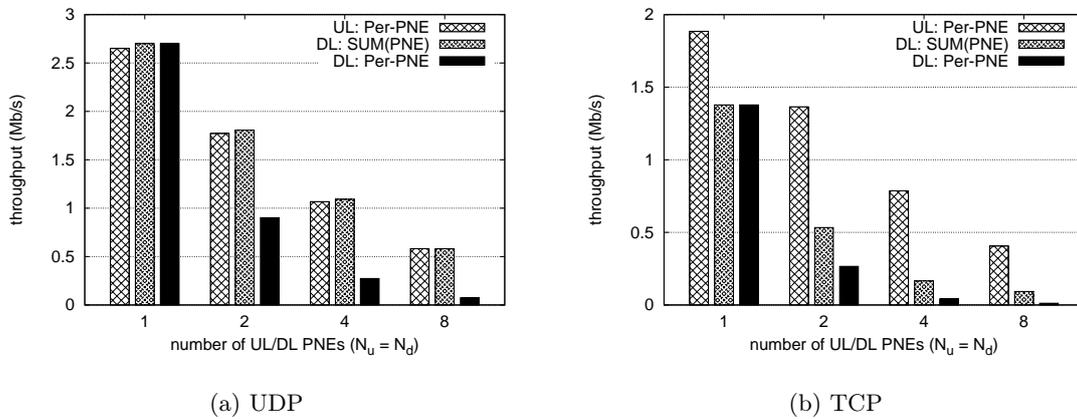


Figure 3: Per-device throughput with different numbers of UL/DL devices.

3 QoS control mechanism

We propose QoS control mechanism that provides QoS service and BE service with the quantitative assurance for QoS and the relative fairness in terms of throughput, respectively. We describe the overall structure and signaling of the proposed mechanism in the first subsection, and the rate control and admission control, two main components of the proposed mechanism, in the subsequent subsections. Lastly, we discuss several issues to elaborate and extend the proposed mechanism.

3.1 Overall structure and signaling

Figure 4 shows the operational flow chart of the proposed QoS control mechanism. To establish a service flow (SF) in advance of service start, a PNE transmits a *service request* (S-REQ) message to the PN GW, where several properties of the service (e.g., required data rate, packet size, and protocol) are described. If the required data rate is not specified in the S-REQ message, we consider such a service as BE service, otherwise the service is considered as QoS service. On receiving the service request, the PN GW classifies the service as QoS service or BE service. For QoS service, the admission control is applied so that the service can be admitted only if its required data rate can be assured. For BE service, the service is admitted without admission

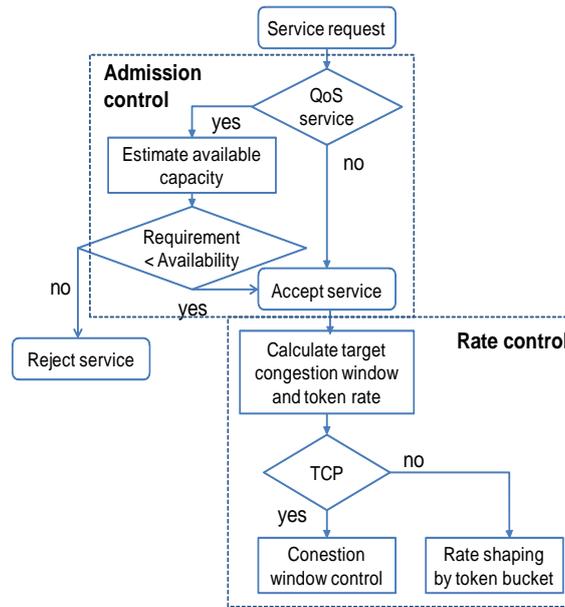


Figure 4: Operational flow chart of the proposed QoS control mechanism.

control and the remaining resource is evenly allocated to BE services under the rationale that the priority of QoS service is strictly higher than that of BE service. Once the service is admitted, the PN GW broadcasts a *service response* (S-RSP) message to the PNEs, to acknowledge the resource reservation for the QoS service and to inform the PNEs with the on-going BE services of the change in the available resource. The S-RSP message includes target values of TCP congestion window (CGW) and token generation rate for both QoS and BE services, which are calculated by the PN GW based on the QoS requirement, available channel capacity, and the number of BE SFs. The rate control is implemented by congestion window control for TCP flows or token bucket algorithm for UDP flows. Receiving S-RSP message, the PNE sets the advertised window size (i.e., the maximum congestion window size) for the TCP flow, as the value in the S-RSP message, to control its transmission rate. For the UDP flow, PN GW or PNE sets the token generation rate for the flow as the value in the S-RSP message, so that its transmission rate can be regulated around the target rate. When a service is terminated, a *service termination* (S-END) message is delivered to the PN GW so as to release the resource occupied by the terminating service and redistribute it to the PNEs with existing BE services. The PN GW recalculates the target values of congestion window and token generation rate and informs PNEs of the updated values by broadcasting the S-RSP message.

3.2 Rate control

We propose two mechanisms for rate control; congestion window control for TCP flows and rate shaping for UDP flows. First, we focus on the rate control for TCP flows. It is shown from [13] that by controlling the size of congestion window, the packet loss due to buffer-overflow in the AP's buffer can be avoided and the unfair channel sharing among TCP flows can be alleviated. We extend this idea to QoS service and BE service delivered by TCP flows. To assure the resource reservation for QoS service and fair resource allocation among BE services, we determine the value of the maximum congestion window for service i , denoted as $CGW_{max,i}$, as;

$$CGW_{max,i} = w_i B, \quad (3)$$

where

$$w_i = \begin{cases} \frac{R_{QoS,i}}{\tilde{C}} & \text{for QoS service,} \\ \left(1 - \frac{\sum_j^{N_{QoS}} R_{QoS,j}}{\tilde{C}}\right) \frac{1}{N_{BE}} & \text{for BE service.} \end{cases} \quad (4)$$

Here, N_{QoS} and N_{BE} are the number of QoS and BE service flows, respectively, $R_{QoS,i}$ denotes the required data rate for the i th QoS service, and \tilde{C} is the estimated capacity of WLAN. Note that the size of CGW limits the number of on-the-fly packets, i.e., the number of packets that can be transmitted before receiving TCP ACK. From (3) and (4), we can see that a dedicated portion of buffer is reserved for each QoS service for assuring its desired data rate and the remaining portion of buffer is evenly distributed to BE services for fairness.

Proposition 1. *The CGW control in (3) and (4) prevents packet loss due to buffer-overflow in the AP.*

Proof: From (3) and (4), it can be shown that

$$\sum_i^{N_{QoS}+N_{BE}} CGW_{max,i} = B. \quad (5)$$

Since the number of on-the-fly packets is limited by the CGW and the packets may be stored in the buffer except the AP, they are not dropped in the AP's downlink buffer due to buffer-overflow. Therefore, the proposed CGW control can remove one cause of unfair channel sharing, i.e., asymmetric congestion control in response of packet loss. \square

Proposition 2. *Assume that the wireless link is a bottleneck link so that the queuing delay in the downlink buffer of AP is a dominant factor of round trip time (RTT) and that the packet loss rate in the wireless link is negligible due to link-layer collision avoidance and retransmission mechanisms. Then, the CGW control in (3) and (4) assures the desired data rate for QoS service and achieves fair and full channel utilization for BE service.*

Proof: From Proposition 1 and the assumptions of Proposition 2, we can consider that the packet loss, which may occur due to buffer-overflow, collision among contending stations, and the poor quality of wireless channel, is insignificant. Then, the size of congestion window increases up to its maximum value according to TCP congestion control, and the steady-state throughput for the i th service flow, th_i^* , becomes

$$th_i^* = \frac{CGW_i}{T_{RTT,i}^*}, \quad (6)$$

where $T_{RTT,i}^*$ is the value of RTT in steady state for the i th flow. Decomposing $T_{RTT,i}^*$ as queuing delay in the AP's downlink buffer ($T_{q,i}$) and the other delay ($T_{o,i}$) (e.g., processing delay at a node and link propagation delay), $T_{RTT,i}^*$ can be represented as

$$T_{RTT,i}^* = T_{q,i} + T_{o,i} \approx \frac{B}{\tilde{C}}, \quad (7)$$

from the assumptions of Proposition 2, i.e., $T_{q,i} \approx B/\tilde{C} \gg T_{o,i}$, where \tilde{C} is the effective capacity of WLAN in steady state. From (3) and (7), we can rewrite (6) as

$$\begin{aligned} th_i^* &\approx \frac{w_i B}{B/\tilde{C}} = w_i \tilde{C} \\ &= \begin{cases} R_{QoS,i} & \text{for QoS service,} \\ \left(\tilde{C} - \sum_j^{N_{QoS}} R_{QoS,j}\right) \frac{1}{N_{BE}} & \text{for BE service.} \end{cases} \end{aligned} \quad (8)$$

Also, the total throughput in steady state becomes from (8)

$$\sum_i^{N_{QoS}+N_{BE}} th_i^* = \tilde{C}. \quad (9)$$

Consequently, it is shown from (8) and (9) that the proposed CGW control mechanism assures the desired data rate for QoS service and allows BE services to fully share the channel in a fair manner. \square

Next, we focus on the rate shaping for UDP flows. Whenever a new service is requested or terminated, the PN GW calculates the token rate (R_{token}) for each service as the value in the rightmost of (8) and broadcasts this value to the PNEs via **S-RSP** message. Receiving the **S-RSP** message, the UL PNE sets the token generation rate for UL UDP flow as the value received. Similarly, the PN GW sets the token generation rate for DL UDP flows. When a packet arrives, the corresponding amount of tokens are removed from the bucket and the packet is served. If the remaining token is less than the size of a new arriving packet, its service is deferred until the amount of tokens becomes larger than the packet size. In this way, the transmission rate of UDP flow can be regulated around the target rate and the proposed mechanism still works even when TCP and UDP flows coexist.

3.3 Admission control

Consider that a new QoS service requests the data rate of $R_{QoS,i}$ and there are N_{QoS} existing QoS services. The new QoS service will be admitted only if the following condition is satisfied,

$$R_{QoS,i} < \left(\alpha \tilde{C} - \sum_j^{N_{QoS}} R_{QoS,j} \right) \quad (10)$$

where $\alpha(0 < \alpha < 1)$ is a control parameter.

The key point of the proposed admission control is estimating the capacity accurately, which is described below. Consider that a PNE sends or receives l -byte (excluding TCP/IP header) TCP data packet to or from the PN GW at the PHY rate of r Mb/s according to the channel access mechanism of IEEE 802.11 DCF. Let us define $T_{data}(r, l)$ and $T_{ack}(r)$ as the average time in the unit of μ s required to successfully serve the TCP data packet and TCP ACK packet, respectively; they can be represented as

$$\begin{aligned} T_{data}(r, l) &= t_{DIFS} + \overline{t_{bo}} + t_{MSDU}(r, l) + t_{SIFS} + t_{MACK} \\ T_{ack}(r) &= t_{DIFS} + \overline{t_{bo}} + t_{MSDU}(r, 0) + t_{SIFS} + t_{MACK}. \end{aligned} \quad (11)$$

In (11), t_{DIFS} and t_{SIFS} are DCF inter frame space (DIFS) and short inter frame space (SIFS), respectively, t_{MACK} is the time required to send MAC-layer ACK frame, which is constant, $\overline{t_{bo}}$ is the average backoff time before sending a MAC frame. Moreover, $t_{MSDU}(r, l)$ is the time required to send a MAC service data unit (MSDU) at the rate of r Mb/s, which can be given as;

$$t_{MSDU}(r, l) = t_{PHY} + \frac{(l + l_h)8}{r}, \quad (12)$$

where t_{PHY} is the time required to send PHY layer preamble and header and l_h is the size of TCP/IP/MAC header in byte. Note that r is selected among pre-defined rates (e.g., 11, 5.5, 2, and 1 Mb/s for 802.11b) depending the channel quality and the link adaptation mechanism adopted.

Next, we calculate $\overline{t_{bo}}$. Let us define p_c as the probability of packet collision in steady state, and $CTW(n)$ as the size of contention window under the condition that the transmission succeeds after n consecutive failures. On the detection of transmission failure, the contention window size is doubled according to the binary exponential backoff (BEB) mechanism of IEEE 802.11, it can increase up to the maximum value of CTW_{max} and returns to the minimum value of CTW_{min} if the transmission succeeds, i.e.,

$$CTW(n) = \begin{cases} 2^n CTW_{min} & \text{for } n \leq m, \\ CTW_{max} & \text{for } n > m. \end{cases} \quad (13)$$

where $m = \log_2(CTW_{max}/CTW_{min})$. Then, \overline{CTW} becomes³

$$\begin{aligned} \overline{CTW} &= \sum_{n=0}^m p_c^n (1-p_c) 2^n CTW_{min} + \sum_{n=m+1}^{\infty} p_c^n (1-p) CTW_{max} \\ &= CTW_{min} \frac{1-p_c - p_c(2p_c)^m}{1-2p_c}. \end{aligned} \quad (14)$$

On the other hand, the probability of packet collision, p_c can be represented with \overline{CTW} . According to the *p-persistent* model [17], which closely approximates the standard IEEE 802.11 protocol, each node makes a transmission attempt with the probability of $p_a = 2/\overline{CTW}$ in steady state. Then, the probability that a station collides with any of the other stations becomes

$$p_c = 1 - (1-p_a)^{M-1} = 1 - (1 - 2/\overline{CTW})^{M-1}, \quad (15)$$

where M is the number of active contending stations. We can get p_c and \overline{CTW} by solving two equations (14) and (15) numerically⁴. A station selects a random backoff counter uniformly distributed between zero and $\overline{CTW}-1$. If a station decrements its backoff counter in the backoff stage, the other $(M-1)$ active contending stations also decrement their own backoff counters. Therefore, the effective backoff time can be approximated as

$$\overline{t_{bo}} = \frac{\overline{CTW} - 1}{2} \frac{1}{M} t_{slot}, \quad (16)$$

where t_{slot} is the slot time. From (11) – (16), $T_{data}(r, l)$ and $T_{ack}(r)$ are obtained. Finally, we can get the capacity of WLAN, $C(r, l)$ (Mb/s), when stations send/receive l -byte TCP data packet at the rate of r Mb/s;

$$C(r, l) = \frac{8l}{T_{data}(r, l) + T_{ack}(r)}. \quad (17)$$

Table 2 lists several parameters of IEEE 802.11b used in the analysis model.

We can extend the analytical capacity model to the case where the link adaptation mechanism is applied so that the PHY-layer transmission rate changes depending on the channel quality. When the auto rate fallback (ARF) algorithm [18], which is the most common link adaptation algorithm, is used, the distribution of PHY-layer transmission rate can be obtained from the analysis result in [19]. The rate distribution can also be obtained by measuring signal to noise ratio (SNR) experimentally when a close-loop SNR-based link adaptation algorithm is used. Let

³Here, we assume that the number of retransmissions is infinite even though it is limited to a certain value in the IEEE 802.11 standard. However, this assumption is not invalid since the probability that the number of retransmissions reaches the maximum value is very low.

⁴The simulation study in Section 4.1 shows that the actual capacity under the proposed mechanism is little deviated from the analytical capacity derived with the condition of $M = 1$, i.e., $p_c = 0$ from (15).

Table 2: Parameters of IEEE 802.11 in the analysis model.

parameter	value
slot time (t_{slot})	20 μ s
SIFS (t_{SIFS})	10 μ s
DIFS (t_{DIFS})	50 μ s
ACK (t_{MACK})	248 μ s
PHY header (t_{PHY})	192 μ s
TCP/IP/MAC header (l_h)	74 byte
minimum contention window (CTW_{min})	32
maximum contention window (CTW_{max})	1024

us define $p(r_i)$ as the probability that the PHY-layer rate is $r_i \in \mathbb{R}$. The effective capacity can be obtained as

$$\tilde{C}(l) = \sum_{r_i \in \mathbb{R}} C(r_i, l) p(r_i). \quad (18)$$

3.4 Discussion

Adjustment of the required data for admission control

The proposed admission control makes the decision on the admission of QoS service by comparing the required data rate and the available capacity, as shown in (10). However, the total capacity, \tilde{C} , is not constant but depends on several factors such as packet size, transport-layer protocol, number of contending nodes, and channel quality. For example, the capacity of WLAN decreases as the packet size decreases, because the MAC/PHY overhead required to transmit a frame becomes relatively large as the packet size decreases. Recall that the analytical capacity in Section 3.3 is derived under the condition that the stations send/receive l -byte TCP packets. Also note that the maximum size of Ethernet frame is 1500 bytes and the size of VoIP packet is typically a few hundreds bytes, and that the data service is usually served with TCP while some real-time service may be served with UDP. We need to cope with the variation of capacity due to difference in the packet size, transport-layer protocol, and channel quality. For this purpose, we propose an approach of adjusting the required data rate of QoS service, instead of adjusting the total capacity.

Consider that a QoS service is requested with the required data rate of $R_{QoS,i}$ and the average TCP packet size of l byte, which is different from the standard packet size of l_o used in the capacity estimation. Also, consider that the transmission rate for the service is r Mb/s⁵. In order to deal with the capacity variation due to different packet size, the required data rate for the QoS service with the packet size of l is adjusted to the equivalent data rate, $R'_{QoS,i}$:

$$R'_{QoS,i} = \frac{\tilde{C}(l_o)}{C(r, l)} R_{QoS,i}. \quad (19)$$

From (19), $R'_{QoS,i}$ increases as the packet size decreases to compensate for the decrease of capacity. Similarly, $R'_{QoS,i}$ decreases if the channel quality for the service is relatively good because the time required to send a packet with the higher transmission rate is smaller than that with the lower transmission rate, i.e., a station in good channel quality consumes less resource to serve a

⁵The transmission rate may change during service time due to varying channel quality. However, the channel quality is mostly depends on the path-loss, which is determined by the distance between transmitter and receiver; and thus the transmission rate does not change severely as long as the station does not move.

packet. Note that the adjusted rate requirement, $R'_{QoS,i}$ is used not only in the admission control (10), but also in the TCP window control (4).

In the same way, the required data rate for a QoS service with UDP can be adjusted. In contrast to TCP, UDP does not require transport-layer ACK transmission from the receiver, resulting in less overhead and higher capacity. The capacity with UDP traffic, denoted as $C_{udp}(r, l)$ can be estimated by replacing (17) with (20);

$$C_{udp}(r, l) = \frac{8l}{T_{data}(r, l)}. \quad (20)$$

Then, the equivalent required data rate for a UDP QoS service, $R'_{QoS,i}$, becomes

$$R'_{QoS,i} = \frac{\tilde{C}(l_o)}{C_{udp}(r, l)} R_{QoS,i}. \quad (21)$$

Extension to support weighted fairness for BE services

The proposed scheme can be easily extended to support weighted fairness for BE services. Consider that a BE service has the corresponding service weight $w_{BE} (\geq 1)$, whose value is integer and determined by a service operator of CPNS depending on the relative service priority. Also, we consider the service weight is delivered to the PN GW in the **S-REQ** message. Then, the PN GW calculates the maximum congestion window (CGW_{max}) for TCP BE service and the token generation rate (R_{token}) for UDP BE service as

$$\begin{aligned} CGW_{max} &= \left(1 - \frac{\sum_j^{N_{QoS}} R_{QoS,j}}{\tilde{C}} \right) \frac{B}{\sum_k^{N_{BE}} w_{BE,k}}, \\ R_{token} &= \left(\tilde{C} - \sum_j^{N_{QoS}} R_{QoS,j} \right) \frac{1}{\sum_k^{N_{BE}} w_{BE,k}}. \end{aligned} \quad (22)$$

These values of CGW_{max} and R_{token} are delivered to PNEs in the **S-RSP** message. Receiving the **S-RSP** message, each PNE sets the congestion window size and token generation rate for its BE service that has the service weight of $w_{BE,k}$, each of which is denoted as $CGW'_{max,k}$ and $R'_{token,k}$, respectively, as;

$$\begin{aligned} CGW'_{max,k} &= w_{BE,k} CGW_{max}, \\ R'_{token,k} &= w_{BE,k} R_{token}. \end{aligned} \quad (23)$$

Note that $CGW'_{max,k} = CGW_{max}$ and $R'_{token,k} = R_{token}$ for a normal BE service with $w_{BE,k} = 1$. From (22) and (23), we observe that BE services share the buffer and capacity in a weighted manner and this extension for weighted service does not degrade QoS service since

$$\begin{aligned} \sum_k^{N_{BE}} CGW'_{max,k} &= \left(1 - \frac{\sum_j^{N_{QoS}} R_{QoS,j}}{\tilde{C}} \right) B < B, \\ \sum_k^{N_{BE}} R'_{token,k} &= \tilde{C} - \sum_j^{N_{QoS}} R_{QoS,j} < \tilde{C}. \end{aligned} \quad (24)$$

Further service differentiation for QoS

The proposed mechanism can further differentiate service to enhance QoS. The 802.11e EDCA can be straightforwardly incorporated with the proposed mechanism so that the channel access for the QoS service can be made preferentially over the BE service. Moreover, such service differentiation can be simply realized without employing 802.11e EDCA. There is an interface queue between IP layer and MAC layer, which typically operates in a first-come-first-serve (FCFS) manner. By implementing dual queue, one for QoS service and the other for BE service, and adopting the priority scheduling algorithm for the dual queue, the packets of BE service is not served as long as there is the packets of QoS service waiting for the service. In this way, the delay for QoS service can be minimized.

4 Performance evaluation

In this section, we perform the extensive ns-2 simulations to evaluate the performance of the proposed QoS control mechanism in terms of QoS assurance and fair channel sharing. The simulation configuration is same to that in Section 2 and 802.11 MAC/PHY parameters used in the simulations are listed in Tab. 2.

4.1 Validation for capacity estimation

First, we validate the methodology for capacity estimation by comparing the simulation results with the analysis results. The capacity is affected by the packet size or transport-layer protocol as described in Section 3.3, and the estimated capacity is used in the admission control and rate control; therefore the accurate estimation of capacity plays a key role in the proposed mechanism. We consider the following two scenarios;

- TCP-only: We observe the network capacity with different TCP packet size ($l = 500$ and 1500 bytes) and different number of UL and DL PNEs ($N_u = N_d = 1 \sim 5$).
- Mixed: In this scenario, each two PNEs upload and download 1 Kbyte TCP packets, while each N_{udp} PNEs upload and download UDP packets, respectively. The size and interval of UDP packet are set to 1500 bytes and 30 ms (i.e., 360 Kb/s), and 200 bytes and 20 ms (i.e., 32 Kb/s), each of which is set to consider the real-time video and audio service for mobile device, respectively.

Fig. 5 compares the capacity obtained from simulation with that obtained from analysis, for the above two cases. From Fig. 5(a), we observe that the simulation result is almost immune to the number of PNEs. This result seems not to agree with the well-known WLAN capacity model [17], [20], where the capacity mostly decreases as the number of stations increases due to the increased collision probability. These previous models make an assumption of saturated traffic, i.e., every station always has packets to send and it always participates in the MAC-layer contention. However, this explanation does not hold in our mechanism because the admission control along with the rate control assures the sum of transmission rates for all the service is regulated not to exceed the available capacity so that every station cannot always take part in the MAC-layer contention. Note that the analysis result shown in Fig. 5 is derived under the condition that the number of active contending stations is one, i.e., $M = 1$ and $p_c = 0$ accordingly from (15), resulting in the constant capacity, regardless of N_d and N_u . On the other hand, as was expected, the capacity decreases with the smaller size of packet. However, Fig. 5(b) shows that the capacity linearly increases or decreases with respect to the number of PNEs with UDP traffic in the mixed traffic case. From the analysis in Section 3.3, the estimated capacity in the

case of 1500-byte and 200-byte UDP packet is 5.00 and 1.57 Mb/s, respectively, and that in the case of 1-Kbyte TCP packet is 3.22 Mb/s.

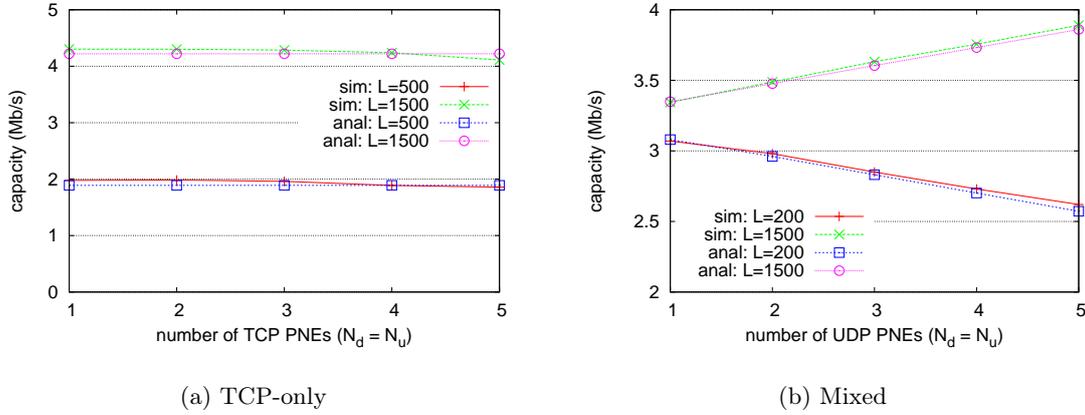


Figure 5: Comparison of capacity obtained from simulation and capacity derived from analysis.

Fig. 5 shows that there is little difference in the simulation results and analysis results for both cases. It is important to note that the capacity (\tilde{C}) with the standard packet size of 1KB is used as decision criteria of admission control (see (10)), and the required data rate is scaled if the packet size or transport-layer protocol is different from those used to estimate the capacity as shown in (19) and (21). The results in Fig. 5 confirm that the estimated capacity is accurate, regardless of packet size and/or transport-layer protocol, and that the capacity can be easily estimated under the assumption of negligible collision probability, regardless of the number of stations. Hereafter, we set the packet size of QoS and BE service as 1 Kbyte to concentrate on evaluating performance without the effect of packet size.

4.2 QoS assurance and fair channel sharing

The objective of this simulation is to evaluate the performance of the proposed mechanism in terms of QoS assurance and fair channel sharing. Figure 6(a) shows per-PNE throughput under the simulation scenario in Section 2.1 (see Tab. 1). By comparing the case without QoS control (see Fig. 2), we observe that the proposed mechanism strictly assures the desired data rate for QoS services, e.g, 1Mb/s for IPTV (DL) and 1Mb/s for surveillance camera (UL), and the assurance is hardly affected by other services. The proposed mechanism completely removes the unfairness between DL and UL BE services; the throughputs of the DL mobile phone and UL camcorder are almost equal during the whole simulation time when both coexist. Moreover, the severe fluctuation in the throughput is significantly reduced, compared to the case without QoS control. Table 3 compares the per-PNE throughput with the proposed QoS control and that without QoS control. During $t = [50-100]$ s and $[100-150]$ s, the fairness index is increased from 0.53 and 0.45 to 1.01 and 1.00, respectively, due to the proposed QoS control mechanism. Also, the total throughput of the proposed mechanism is not decreased to assure QoS and fairness, rather it is slightly increased in some cases, compared to the case without QoS control.

Next, we focus on the delay of streaming service (IPTV) for three cases;

- BASE: No QoS control mechanism is implemented.
- QoS: The proposed QoS control mechanism is implemented and the PN GW serves packets with the FCFS scheduling discipline.

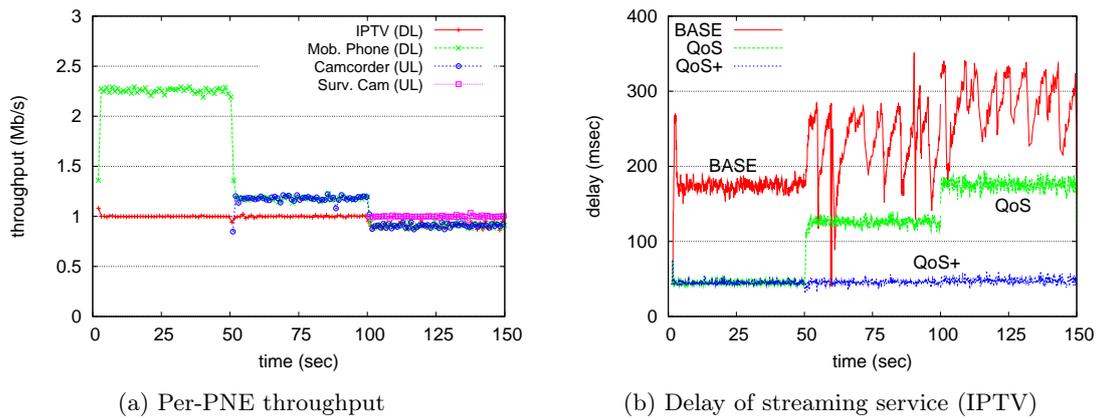


Figure 6: Throughput and delay with the proposed QoS control mechanism for the simulation scenario in Tab. 1.

Table 3: Comparison of per-PNE throughput in each time interval.

interval (sec)	device	direction	throughput (Mb/s)	
			w/o control	with control
0 ~ 50	IPTV	DL	1.00	1.00
	Mob. phone	DL	2.25	2.24
50 ~ 100	IPTV	DL	0.78	1.00
	Mob. phone	DL	0.87	1.18
	Camcorder	UL	1.63	1.17
100 ~ 150	IPTV	DL	0.51	1.00
	Mob. phone	DL	0.71	0.91
	Camcorder	UL	1.58	0.91
	Surv. Cam	UL	1.00	1.00

- **QoS+**: This enhances the proposed QoS control mechanism by implementing dual queue in the PN GW and applying the strict priority scheduling discipline to the dual queue, as discussed in Section 3.4.

As shown in Fig. 6(b), the delay of **BASE** is higher than that of **QoS+** up to 7.5 times and it oscillates highly after a UL service starts at $t = 50$ s, while the delay of **QoS+** is maintained at the smallest value (about 45 ms) and it does not increase due to other services. In the case of **QoS**, the variation of delay is small as **QoS+**, but the delay increases once other service starts; the average delay is about 45, 125, and 170 ms, in the time interval of [0-50]s, [50-100]s, and [100-150]s, respectively. We also measure the jitter of streaming service, defined as standard deviation of delay⁶. In the case of **BASE**, the jitter increases from 16 ms to 77 ms in the time interval of [0-50]s and [100-150]s, respectively. The jitter of **QoS** during $t = [100-150]$ s increases from 3.64 ms to 7.13ms by about 2 times, compared to the time interval of $t = [0-50]$ s. However, as can be expected from 6(b), the jitter of **QoS+** is not affected by other service and is maintained below 3.9 ms for the whole simulation time, reconfirming that the proposed mechanism can minimize delay of **QoS** service, as well as assuring its desired data rate.

⁶The jitter can be defined in several ways to quantify the variation in the delay of successive packets. It is an important performance index for streaming service, as well as delay.

Table 4: Fairness and capacity with different number of devices.

(N_u, N_d)	(1,1)	(1,2)	(1,4)	(2,1)	(2,2)	(2,4)	(4,1)	(4,2)	(4,4)
fairness index	0.987	0.988	1.004	1.000	1.014	1.000	0.992	0.999	0.982
capacity (Mb/s)	3.27	3.35	3.34	3.26	3.32	3.34	3.31	3.26	3.27

4.3 Fairness with different number of devices, weights, and transmission rates

Here, we focus on evaluating the performance of the proposed mechanism from the viewpoints of fairness and capacity for TCP BE services. Table 4 shows fairness index and total throughput with different number of UL/DL devices, N_u and N_d . The fairness index is not much deviated from the ideal value of one, it ranges between 0.982 and 1.014 in all the cases of N_d and N_u . Also, the capacity is little affected by the number of devices, it ranges between 3.26 Mb/s and 3.35 Mb/s and close to the analysis result of 3.23 Mb/s. These results validate that the proposed algorithm assures fair channel sharing among BE services without debasing the total throughput, regardless of the number of devices.

The next simulation evaluates how the proposed mechanism can support the weighted fairness for BE services, which is discussed in Section 3.4. Figure 7 shows throughput for six PNEs (three DL PNEs and another three UL PNEs) that have different service weights of 1, 2, or 4. We observe that the throughput is almost in proportion to the service weight, the throughput of PNE5/PNE6 is higher than that of PNE1/PNE2 by 3.96 times and that of PNE3/PNE4 is higher than that of PNE1/PNE2 by 1.97 times. Moreover, there is no notable difference between throughput of DL PNE and UL PNE that have the same service weight. Since the proposed mechanism assures weighted fairness, the CPNS operator can assign different service weight based on the service priority, service fee, or operational policy, to increase its revenue.

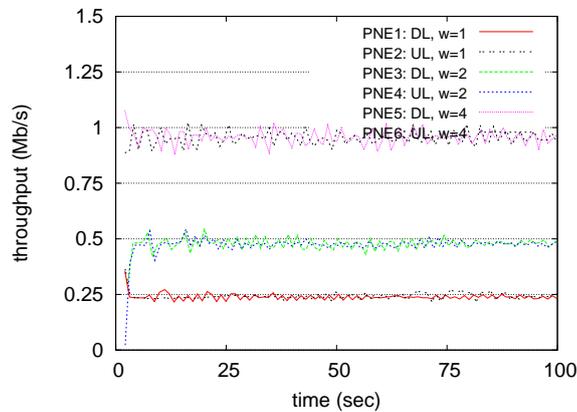


Figure 7: Per-PNE throughput with different service weights.

Finally, we observe the performance of fairness in multi-rate WLAN configuration. Table 5 shows per-PNE throughput, total throughput, and fairness index, as well as indicating the direction and transmission rate (HI and LO denote the transmission rate of 11 Mb/s and 2 Mb/s, respectively.). The difference in the per-PNE throughput is insignificant, i.e., the fairness index is almost one, regardless of transmission direction and transmission rate. However, the total throughput decreases as the number of low-rate PNEs increases, because the low-rate PNE occupies the channel longer than the high-rate PNE to transmit a packet. The proposed mechanism assures *throughput-fairness*, regardless of transmission rate of PNEs. Moreover, it is expected from the results of Fig. 7 and Tab. 5 that the proposed mechanism can be extended

Table 5: Fairness and capacity with different transmission rates.

Per-PNE throughput (Mb/s)				capacity (Mb/s)	fairness index
DL: HI	UL: HI	DL: LO	UL: LO		
1.626	1.645	-	-	3.271	0.988
0.863	0.880	0.871	-	2.614	0.985
0.574	0.582	0.561	0.558	2.275	0.996

to support *time-fairness*, i.e., allocate channel occupation time fairly, by assigning high weight to the high-rate PNE, to improve the total throughput without impairing fairness.

5 Conclusion

In this paper, we proposed a generalized framework to assure QoS and fairness for WLAN-based CPNS. We identified the problem of QoS degradation and unfair channel sharing that occurs in the existing WLAN-based CPNS, and analyzed its cause from the viewpoints of unfair channel access of MAC and asymmetric congestion control of TCP. By employing admission control and rate control, the proposed mechanism provides absolute guarantee for QoS service in terms of data rate, at the same time, it provides weighted fairness for BE service without decreasing the efficiency of channel sharing. Moreover, the assurance of QoS and fairness is effective, regardless of service direction, packet size, transmission rate, and transport-layer protocol. Through analysis and simulation, it was confirmed that the proposed mechanism effectively achieve QoS assurance for QoS service and fair channel access for BE service.

Acknowledgement

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (KRF), funded by the Ministry of Education, Science, and Technology [Grant No. 2010-0008711], and by Information Technology Research Center (ITRC) support program through the National IT Industry Promotion Agency (NIPA), funded by the Ministry of Knowledge Economy, Korea [Grant No. NIPA-2011-C1090-1121-0005]. Corresponding author is G.-M. Jeong.

Bibliography

- [1] Open Mobile Alliance, Converged Personal Network Service Architecture Candidate Version 1.0, OMA-AD-CPNS-V1_0-20100615-C, Jun 2010.
- [2] Open Mobile Alliance, Converged Personal Network Service Requirements, Candidate Version 1.0, OMA-RD-CPNS-V1_0-20091117-C, Nov. 2009.
- [3] IEEE 802.11 Working Group, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification, IEEE Std 802.11-2007, June 2007.
- [4] The Network Simulator, ns-2, <http://www.isi.edu/nsnam/ns/>
- [5] IEEE 802.11 Working Group, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specification, Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements IEEE Std 802.11e-2005, Nov. 2005.

-
- [6] I. Aad and C. Castelluccia, Differentiation Mechanisms for IEEE 802.11, in *Proc. of IEEE INFOCOM*, 209-218, 2001.
- [7] Q. Ni, L. Romdhani, and T. Turetli, A Survey of QoS Enhancements for IEEE 802.11 Wireless LAN, *Wiley Journal of Wireless Communication and Mobile Computing*, 4(5):547-566, 2004.
- [8] H. Zhu, M. Li, I. Chlamtac, and B. Prabhakaran, A survey of Quality of Service in IEEE 802.11 Networks, *IEEE Wireless Communications*, 11(4):6-14, 2004.
- [9] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan, Achieving MAC Layer Fairness in Wireless Packet Networks, in *Proc. of ACM MobiCom*, 87-98, 2000.
- [10] D. Qiao and K.G. Shin, Achieving Efficient Channel Utilization and Weighted Fairness for Data Communications in IEEE 802 WLAN under the DCF, in *Proc. of Int. Workshop Quality of Service (IWQoS)*, 227-236, 2002.
- [11] M. Bottigliengo, C. Casetti, C.-F. Chiasserini, and M. Meo, Smart Traffic Scheduling in 802.11 WLANs with Access Point, in *Proc. IEEE Vehicular Technology Conf.-Fall (VTC03)*, 2227-2231, 2003.
- [12] S. Kim, B.-S. Kim, and Y. Fang, Downlink and Uplink Resource Allocation in IEEE 802.11 Wireless LANs, *IEEE Trans. Vehicular Technology*, vol. 54, no. 1, pp. 320-327, Jan. 2005.
- [13] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, Understanding TCP Fairness over Wireless LAN, in *Proc. of IEEE INFOCOM*, pp. 863-872, 2003.
- [14] Y. Wu, Z. Niu, and J. Zheng, Study of the TCP Upstream/Downstream Unfairness Issue with Per-Flow Queuing over Infrastructure-Mode WLANs, *Wireless Comm. and Mobile Computing*, vol. 5, no. 4, pp. 459-471, 2005.
- [15] D.J. Leith, P. Clifford, D. Malone, and A. Ng, TCP Fairness in 802.11e WLANs, *IEEE Communications Letters*, vol. 9, no. 11, pp. 964-966, 2005.
- [16] S. Choi, K. Park, and C. Kim, Performance Impact of Inter-Layer Dependence in Infrastructure WLANs, *IEEE Trans. Mobile Computing*, vol. 5, no. 7, pp. 829-845, 2006.
- [17] F. Calì, M. Conti, E. Gregori, Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit, *IEEE Trans. on Networking*, vol. 8, no. 6, pp. 785-799, Dec. 2000.
- [18] A. Kamerman and L. Monteban, WaveLAN-II: A High-performance wireless LAN for the unlicensed band, *Bell Lab Technical Journal*, pp. 118-133, 1997.
- [19] J. Choi, K. Park, and C.-K. Kim, Analysis of Cross-Layer Interaction in Multirate 802.11 WLANs, *IEEE Trans. on Mobile Computing*, vol. 8, no. 5, pp. 682-693, May 2009.
- [20] G. Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535-547, March 2000.