摘要

IEEE 802.11e 於媒體存取層定義了一個嶄新的傳送協調機制，名為『混合協調機制』。此機制於了解各個工作站的服務品質需求後，分配傳送機會給工作站。然而，在混合協調機制裡屬於免競爭形態的 HCCA 中，對於傳送機會分配的參考設計僅適用於固定位元速率的訊務，對於變動位元速率的訊務，封包可能遭受嚴重的遺失。在此計畫中，我們提出 1)高效能傳送機會分配演算法、2)工作站封包多工機制、3)允入控制單元，以保證變動位元速率訊務的服務品質需求，例如：延遲限制、封包遺失機率。我們透過定義等效訊務流、整體封包遺失率機率來得到訊務內與訊務間多工增益，並藉此提高頻寬使用效率。再者，分配集合傳輸機會給各個訊務流時，我們也提出並採用加權遺失公平的服務排程概念。電腦模擬的數據顯示，相較於先前研究成果，我們所提出的方法不僅滿足每個訊務流的服務品質需求並且可達到較高的頻寬使用效率。

關鍵字：服務品質、無線區域網路、排程

Abstract

The Medium Access Control (MAC) of IEEE 802.11e defines a novel coordination function, namely, Hybrid Coordination Function (HCF), which allocates Transmission Opportunity (TXOP) to stations taking their quality of service (QoS) requirements into account. However, the reference TXOP allocation scheme of HCF Controlled Channel Access (HCCA), a contention-free channel access function of HCF, is only suitable for constant bit rate (CBR) traffic. For variable bit rate (VBR) traffic, packet loss may occur seriously. In this project, we generalize the reference design with an efficient TXOP allocation algorithm, a multiplexing mechanism, and the associated admission control unit to guarantee QoS for VBR flows with different delay bound and packet loss probability requirements. We define equivalent flows and aggregate packet loss probability to take advantage of both intra-flow and inter-flow multiplexing gains so that high bandwidth efficiency can be achieved. Moreover, the concept of weighted-loss fair service scheduling is adopted to allocate the aggregate TXOP to individual flows. From numerical results obtained by computer simulations, we found that our proposed scheme meets QoS requirements and results in much higher bandwidth efficiency than previous algorithms.

Index Term: Quality of Service, Wireless LAN, Scheduling

Content

1. Introduction

Wireless networks such as IEEE 802.11 WLANs [1] have recently been deployed widely with rapidly increasing users all over the world. As real-time applications such as VoIP and streaming video are getting more common in daily life, quality of service (QoS) guarantee over wireless networks is becoming an important issue. Generally speaking, QoS support includes guarantee of maximum packet delay, delay jitter, and packet loss probability. To cope with this problem, a new enhancement of WLANs, called IEEE 802.11e [2], was introduced to provide QoS support for real-time traffic. Because of the importance of QoS support in WLANs, IEEE 802.11e had been combined into IEEE 802.11 in 2007 [1].

The QoS-aware coordination function proposed in IEEE 802.11e is called Hybrid Coordination Function (HCF). This function consists of two channel access mechanisms. One is contention-based Enhanced Distributed Channel Access (EDCA) and the other is contention-free HCF Controlled Channel Access (HCCA). The contention-free nature makes HCCA a better choice for QoS support than EDCA [3].

HCCA requires a centralized QoS-aware coordinator, called Hybrid Coordinator (HC), which has a higher priority than normal QoS-aware stations (QSTAs) in gaining channel control. HC can gain control of the channel after sensing the medium idle for a PCF inter-frame space (PIFS) that is shorter than DCF inter-frame space (DIFS) adopted by QSTAs. After gaining control of the transmission medium, HC polls

QSTAs according to its polling list. In order to be included in HC's polling list, a QSTA needs to negotiate with HC by sending the Add Traffic Stream (ADDTS) frame. In this frame, the QSTA describes the traffic characteristics and the QoS requirements in the Traffic Specification (TSPEC) field. Based on the traffic characteristics and the QoS requirements, HC calculates the scheduled service interval (SI) and transmission opportunity (TXOP) duration for each admitted flow.

Upon receiving a poll, the polled QSTA either responds with QoS-Data if it has packets to send or a QoS-Null frame otherwise. When the TXOP duration of some QSTA ends, HC gains the control of channel again and either sends a QoS-Poll to the next station on its polling list or releases the medium if there is no more QSTA to be polled. Polling, together with TXOP allocation, will be referred to as a scheduling scheme of HCCA in this project.

Scheduling schemes can be classified into two categories, namely, static and dynamic. In a static scheduling scheme, HC allocates the same TXOP duration to a QSTA every time it is polled. QSTAs can be polled periodically. Moreover, the SI is often selected as the minimum of delay bound requirements of all traffic flows. The sample scheduler provided in IEEE 802.11e standard document is a typical example of static scheduling scheme. The HC of the sample scheduler allocates TXOP duration based on mean data rate and nominal MAC service data unit (MSDU) size. It performs well for constant bit rate (CBR) traffic. For variable bit rate (VBR) traffic, packet loss may occur seriously. Some other static and periodic scheduling schemes were proposed in [5]–[8]. These schemes generalized the sample scheduler with modified TXOP allocation algorithm and admission control unit so that both delay bound and packet loss probability requirements of admitted traffic flows can be fulfilled. The fact that many real-time VBR applications can tolerate packet loss to certain degree was taken into consideration in those schemes to improve bandwidth efficiency. The sample scheduler does not take advantage of inter-flow multiplexing gain because the TXOP duration allocated to a QSTA is the sum of the TXOP durations allocated to individual flows attached to it. The schemes proposed in [5]–[8] result in higher bandwidth efficiency than the sample scheduler because they considered inter-flow multiplexing effect. In references [5] and [6], it was assumed that all traffic flows have the same delay bound of one SI and the same packet loss probability requirement. In [7], traffic flows are allowed to have different delay bound but identical packet loss probability requirements. A finite buffer is provided for packets with delay bound greater than one SI. With such a buffer, packets that arrived in different previous SIs (and have not violated their delay bound requirement) can share the current TXOP to achieve intra-flow multiplexing gain. The Rate-Variance envelope based Admission Control (RVAC) algorithm presented in [8] uses token buckets for traffic shaping. With the token buckets, the envelope of traffic arrival can be determined. Using the traffic envelope and the given delay bound requirement, one can compute the packet loss probability for an allocated bandwidth. This scheme only considers identical delay bound and packet loss probability requirements.

In contrast to static ones, a dynamic scheduling scheme allocates TXOP duration to a QSTA dynamically, according to system status, to provide delay bound guarantee and/or fairness. Some dynamic scheduling schemes can be found in [9]-[14]. To achieve delay bound guarantee, a dynamic scheduling scheme requires QSTAs to timely report their queue statuses to HC. As an example, in the prediction and optimization-based HCCA (PRO-HCCA) scheme that was proposed recently [11], the SI is set to be smaller than or equal to half of the minimum of delay bounds requested by all traffic flows. As a consequence, compared with static scheduling schemes, QSTAs are polled more frequently, which implies higher overhead for poll frames. Furthermore, static and periodic polling allows QSTAs to easily eliminate overhearing to save energy. Therefore, although dynamic scheduling has the potential to achieve high bandwidth efficiency, it is worthwhile to study static scheduling schemes.

The purpose of this project is to present an efficient static scheduling scheme to provide QoS guarantee for VBR traffic flows with different delay bound and packet loss probability requirements. The proposed scheme achieves both intra-flow and inter-flow multiplexing gains. In this scheme, HC calculates TXOP duration and performs admission control while every QSTA implements a weighted-loss fair service scheduler to determine how the allocated TXOP is shared by traffic flows attached to it. Numerical results obtained by computer simulations show that our proposed TXOP allocation algorithm results in much better performance than previous works. Moreover, the proposed weighted-loss fair service scheduler successfully manages the TXOP so that different delay bound and packet loss probability requirements of all traffic flows can be fulfilled.

The remainder of this project is organized as follows. In Section 2, we describe the system model. In Section 3, we review related previous works. Section 4 contains our proposed TXOP allocation algorithm, the weighted-loss fair service scheduler, and the associated admission control unit. Simulation results are provided and discussed in Section 5. Finally, we draw conclusion in Section 6.

2. System Model

The studied system consists of K QSTAs, called $QSTA_1$, $QSTA_2$, …, and $QSTA_K$ such that $QSTA_i$ has n_i existing VBR flows. Transmission over the wireless medium is divided into SIs and the duration of each SI, denoted by *SI*, is a sub-multiple of the length of a beacon interval T_b. Moreover, an SI is further divided into a contention period and a contention-free period. The HCCA protocol is adopted during contention-free periods.

It is assumed that every QSTA has the capability to measure channel quality to determine a feasible transmission rate which yields a frame error rate sufficiently smaller than the packet loss probability requirements requested by all traffic flows attached to the QSTA. The relationship between measured channel quality and frame error rate can be found in [15].

The QoS requirements of traffic flows are specified with delay bound and packet loss probability. Every QSTA is equipped with sufficiently large buffer so that a packet is dropped if and only if (iff) it violates the delay bound. It is assumed that there are I different packet loss probabilities, represented by P_1, P_2, …, and P_I with $P_1 > P_2 > ... > P_I$, and J possible delay bounds, denoted by D_1, D_2, …, and D_J with $D_1 < D_2 < ... < D_J$. We assume that $D_1 = SI$ and $D_j = \beta_j SI$ for some integer $\beta_j > 1$.

HC allocates TXOPs to QSTAs based on a static and periodic schedule. As illustrated in Fig. 1, the TXOP for $QSTA_k$, denoted by $TXOP_k$, is allocated every SI and is of fixed length. The length of scheduled SI is chosen to be the minimum of all requested delay bounds. Note that SI is updated if a new flow with delay bound smaller than those of existing ones is admitted or the existing flow with the smallest delay bound is disconnected and there is no other existing flow with the same delay bound. In this case, the TXOPs allocated to QSTAs have to be recalculated accordingly.

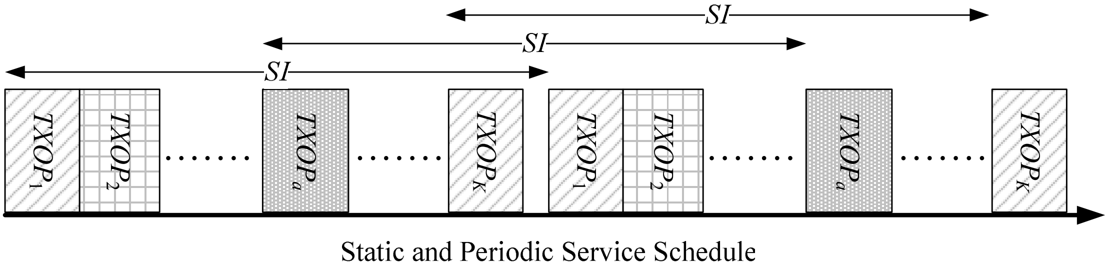

Fig. 1 Static and periodic schedule for 802.11e HCCA.

Consider the existing flows of a specific QSTA, say $QSTA_a$. The n_a flows attached to $QSTA_a$ are classified into groups according to their QoS requirements. Let $F_{i,j}$ represent the set which contains all traffic flows with packet loss probability P_i and delay bound D_j. Furthermore, let $F_i = \bigcup_{1\le j\le J} F_{i,j}$ and $F = \bigcup_{1\le i\le I} F_i$. To reduce computational complexity, we assume that the traffic arrivals of different flows are independent Gaussian processes. Since sum of independent Gaussian random variables remains Gaussian, the aggregated flow of all the flows in set $F_{i,j}$ is Gaussian and will be represented by $f_{i,j}$. For convenience, we shall consider $f_{i,j}$ as a single flow. A separate queue, called $Queue_{i,j}$, is maintained for flow $f_{i,j}$, $1 \le i \le I$ and $1 \le j \le J$. Let $N(\mu_{i,j}, \sigma_{i,j}^2)$ denote the distribution of traffic arrival for flow $f_{i,j}$ in one SI. Note that the values of $\mu_{i,j}$ and $\sigma_{i,j}^2$ can be calculated by $\mu_{i,j} = E(N_{i,j}) \cdot E(X_{i,j})$ and $\sigma_{i,j}^2 = E(N_{i,j}) \cdot VAR(X_{i,j}) + E(X_{i,j})^2 \cdot VAR(N_{i,j})$, where $N_{i,j}$ and $X_{i,j}$ represent, respectively, the number of packets belonging to flow $f_{i,j}$ that arrive in one SI and the packet size. Interested readers can find the derivations in [6].

3. Previous Works

3.1 The Sample Scheduler

Consider $QSTA_a$ which has n_a flows. Let ρ_l, L_l denote, respectively, the mean data rate and the nominal MSDU size of the l^{th} flow attached to $QSTA_a$. HC calculates $TXOP_a$ as follows. Firstly, it decides, for flow l, the average number of packets $\overline{N_l}$ that arrive at the mean data rate during one SI

$$\overline{N_l} = \left\lceil \frac{\rho_l \times SI}{L_l} \right\rceil \tag{1}$$

Secondly, the TXOP duration for this flow is obtained by

$$TD_l = \max\left\{ \overline{N_l} \times \left(\frac{L_l}{R_a^{\min}} + O \right), \frac{L_{\max}}{R_a^{\min}} + O \right\} \tag{2}$$

where $R_a^{\min}$ is the minimum physical transmission rate of $QSTA_a$, and $L_{\max}$ and O denote, respectively, the maximum allowable size of MSDU and per-packet overhead in time units. The overhead O includes the transmission time for an ACK frame, inter-frame space, MAC header, CRC field and PHY PLCP preamble and header.

Finally, the total TXOP duration allocated to $QSTA_a$ is given by

$$TXOP_a = \left(\sum_{l=1}^{n_a} TD_l \right) + SIFS + t_{POLL} \tag{3}$$

where $SIFS$ and t_{POLL} are, respectively, the short inter-frame space and the transmission time of a CF-Poll frame.

Admission control is performed as follows. Assume that $QSTA_a$ negotiates with HC for admission of a new traffic flow, i.e., the $(n_a + 1)^{th}$ flow of $QSTA_a$. For simplicity, we further assume that the delay bound of the new flow is not smaller than SI. The process is similar if this assumption is not true. HC updates *TXOP*$_a$ as $TXOP'_a = TXOP_a + TD_{n_a+1}$. The new flow is admitted iff the following inequality is satisfied

$$\frac{TXOP'_a}{SI} + \sum_{k=1,k\neq a}^{K} \frac{TXOP_k}{SI} \leq \frac{T_b - T_{cp}}{T_b} \tag{4}$$

where T_{cp} is the time used for EDCA traffic during one beacon interval.

It is clear that the TXOP allocation algorithm of the sample scheduler does not consider delay bound and packet loss probability requirements. Moreover, it does not take advantage of inter-flow multiplexing gain.

3.2 Scheme for Traffic Flows with Identical Packet Loss Probability Requirements

In [6], it was assumed that all traffic flows request the same packet loss probability and the same delay bound of one *SI*. The assumption was relaxed in [7] to allow flows requesting different delay bounds but identical packet loss probabilities. We only describe the scheme proposed in [7] because it is a generalization of that presented in [6]. Without loss of generality, assume that the packet loss probability requested by all flows is P_1. As a result, we have $F = F_1$. Further, for ease of description, we assume that there is at least one traffic flow with delay bound D_1.

Again, consider $QSTA_a$ which has n_a flows. The n_a flows are classified into J disjoint sets $F_{1,1}$, $F_{1,2}$, …, and $F_{1,J}$ such that a flow belongs to $F_{1,j}$ iff its delay bound is $\beta_j SI$. Let $f_{1,j}$, $1 \le j \le J$, with traffic arrival distribution $N(\mu_{1,j}, \sigma_{1,j}^2)$ denote the aggregated flow of all the flows in set $F_{1,j}$. The first come first serve (FCFS) service discipline was adopted for packet transmission. The effective bandwidth $c_{1,j}$ of flow $f_{1,j}$ is computed to take advantage of intra-flow multiplexing gain. The effective bandwidth $c_{1,j}$ is defined as the minimum TXOP allocated to flow $f_{1,j}$ to guarantee a packet loss probability smaller than or equal to P_1 for flow $f_{1,j}$. Since the delay bound of flow $f_{1,j}$ is $\beta_j SI$, the effective bandwidth $c_{1,j}$ can be determined with a finite-buffer queueing model where the buffer size is $\beta_j c_{1,j}$, the server transmission capability is $c_{1,j}$, and the desired packet loss probability is P_1. Given the traffic arrival distribution $N(\mu_{1,j}, \sigma_{1,j}^2)$, the effective bandwidth $c_{1,j}$ can be written as $c_{1,j} = \mu_{1,j} + \alpha_{1,j}\sigma_{1,j}$, where $\alpha_{1,j}$ was called the QoS parameter of flow $f_{1,j}$. Derivation of packet loss probability for a finite-buffer system is complicated. Reference [16] provided a good approximation based on the tail probability of an infinite buffer system and the loss probability of a buffer-less system, as shown in equation (5).

$$P_L(x) \approx \frac{P_L(0)}{P(X > 0)} P(X > x) \tag{5}$$

In the above equation, $P_L(x)$ represents the packet loss probability of a finite-buffer system with buffer size *x* and $P(X > x)$ denotes the tail probability above level *x* of an infinite-buffer system. The equation for $P(X > x)$ can be found in [16]. It is pretty complicated and thus is omitted due to space limitation. The equation for $P_L(0)$ can be found in [7] and is given by

$$P_L(0) = Q(\alpha_{1,j}) + \left[\frac{\sigma_{1,j}}{\mu_{1,j}\sqrt{2\pi}} e^{-(\alpha_{1,j}^2/2)} - \left(1 + \frac{\alpha_{1,j}\sigma_{1,j}}{\mu_{1,j}}\right) Q(\alpha_{1,j}) \right] \qquad (6)$$

where $Q(\alpha_{1,j}) = \int_{\alpha_{1,j}}^{\infty} (1/\sqrt{2\pi}) e^{-(x^2/2)} dx$. Having $P(X > x)$, $P(X > 0)$ and $P_L(0)$, one can obtain the (approximate) packet loss probability of a finite-buffer system with server transmission capability $c_{1,j}$ and buffer size $\beta_j c_{1,j}$ as

$$P_L(\beta_j c_{1,j}) \approx \frac{\sigma_{1,j}}{\mu_{1,j}\sqrt{2\pi}} e^{-(\alpha_{1,j}\beta_j c_{1,j}/\sigma_{1,j})} - \frac{\alpha_{1,j}\sigma_{1,j}}{\mu_{1,j}} e^{(\alpha_{1,j}^2/2)-(\alpha_{1,j}\beta_j c_{1,j}/\sigma_{1,j})} Q(\alpha_{1,j}) \qquad (7)$$

Consequently, given mean $\mu_{1,j}$, variance $\sigma_{1,j}^2$, delay bound $\beta_j SI$, and the desired packet loss probability $P_1 = P_L(\beta_j c_{1,j})$, the QoS parameter $\alpha_{1,j}$ can be computed with equation (7) which in turn can be used to derive the effective bandwidth $c_{1,j} = \mu_{1,j} + \alpha_{1,j}\sigma_{1,j}$ for flow $f_{1,j}$.

Let $L_{1,j}$ represent the nominal packet size of flow $f_{1,j}$. The average number of packets which can be transmitted in one SI, denoted by $\overline{N_{1,j}}$, can be estimated as

$$\overline{N_{1,j}} = \left\lceil \frac{c_{1,j}}{L_{1,j}} \right\rceil \qquad (8)$$

The allocated TXOP duration for flow $f_{1,j}$ is given by

$$TD_{1,j} = \max\left\{ \frac{c_{1,j}}{R_a} + \overline{N_{1,j}} \times O \ , \ \frac{L_{\max}}{R_a} + O \right\} \qquad (9)$$

where R_a represents the feasible physical transmission rate of $QSTA_a$.

As mentioned before, using buffer to store packets achieves intra-flow multiplexing gain. To further achieve inter-flow multiplexing gain, an equivalent flow of delay bound D_1, denoted by $\hat{f}_{1,j}$, is defined for flow $f_{1,j}$, $1 \le j \le J$. Let

$N\left(\hat{\mu}_{1,j},\hat{\sigma}_{1,j}^2\right)$ be the traffic arrival distribution of $\hat{f}_{1,j}$. We have $\hat{f}_{1,1}=f_{1,1}$. The equivalent flow $\hat{f}_{1,j}$ for $2\le j\le J$ is obtained by letting its mean and effective bandwidth equal to those of flow $f_{1,j}$, i.e., $\hat{\mu}_{1,j}=\mu_{1,j}$ and $\hat{\alpha}_{1,j}\hat{\sigma}_{1,j}=\alpha_{1,j}\sigma_{1,j}$, where $\hat{\alpha}_{1,j}$ is the QoS parameter of the equivalent flow. Since the delay bound of the equivalent flow $\hat{f}_{1,j}$ is equal to $D_1=SI$, a packet of $\hat{f}_{1,j}$ which arrives in the n^{th} SI will violate its delay bound and be dropped if it is not served in the $(n+1)^{th}$ SI. As a consequence, the effective bandwidth for $\hat{f}_{1,j}$ can be derived based on a buffer-less system. That is, the QoS parameter $\hat{\alpha}_{1,j}$ can be computed according to equation (6) for $P_L(0)=P_1$. It was shown in [7] that $\hat{\alpha}_{1,j}$ can be well approximated by $Q^{-1}\left(P_1\right)$. With the approximation, we have $\hat{\sigma}_{1,j}=\alpha_{1,j}\sigma_{1,j}/Q^{-1}\left(P_1\right)$. After obtaining the equivalent flows $\hat{f}_{1,j}$, $1\le j\le J$, one can determine the aggregate equivalent flow $\hat{f}_1$. Let $N\left(\hat{\mu}_1,\hat{\sigma}_1^2\right)$ denote the distribution of traffic arrival in one SI for the aggregate equivalent flow $\hat{f}_1$. Since sum of independent Gaussian random variables remains Gaussian, we have $\hat{\mu}_1=\mu_{1,1}+\sum_{j=2}^{J}\hat{\mu}_{1,j}$ and $\hat{\sigma}_1^2=\sigma_{1,1}^2+\sum_{j=2}^{J}\hat{\sigma}_{1,j}^2$. Again, given $\hat{\mu}_1$ and $\hat{\sigma}_1^2$, the QoS parameter $\hat{\alpha}_1$ of flow $\hat{f}_1$ can be derived according to equation (6) for $P_L(0)=P_1$. Having $\hat{\alpha}_1$, one can compute the effective bandwidth $\hat{c}_1$ for flow $\hat{f}_1$. The TXOP duration allocated to $QSTA_a$ is then determined as follows

$$TXOP_a=\max\left\{\frac{\hat{c}_1}{R_a}+\overline{N_1}\times O+SIFS+t_{POLL},n_a\times\left(\frac{L_{\max}}{R_a}+O\right)\right\} \tag{10}$$

where

$$\hat{c}_1=\hat{\mu}_1+\hat{\alpha}_1\hat{\sigma}_1 \tag{11}$$

$$\overline{N_1}=\left\lceil\frac{\hat{c}_1}{L_1}\right\rceil \tag{12}$$

In equation (12), $\overline{L_1}$ denotes the weighted average nominal packet size of all the flows in F_1, and is calculated by

$$\overline{L_1} = \frac{\sum_{j=1}^{J} \overline{N_{1,j}} \times L_{1,j}}{\sum_{j=1}^{J} \overline{N_{1,j}}}. \tag{13}$$

The criterion shown in equation (4) was used for admission control.

Clearly, assuming all traffic flows have identical packet loss probabilities is a big constraint of the above scheme. A straightforward solution to handle flows with different packet loss probabilities is to assume that all flows have the most stringent requirement. Unfortunately, such a solution increases the effective bandwidths of flows which allow packet loss probabilities greater than the smallest one. Another possible solution is to compute separately the effective bandwidth $\hat{c}_i$ for aggregated equivalent flow $\hat{f}_i$, $1 \le i \le I$, and allocate $TXOP_a = \sum_{i=1}^{I} \hat{c}_i$. Such a solution, however, does not take advantage of inter-flow multiplexing gain. In the following section, we present our proposed scheme which considers different packet loss probabilities and takes advantage of inter-flow multiplexing gain.

4. Our Proposed Scheme

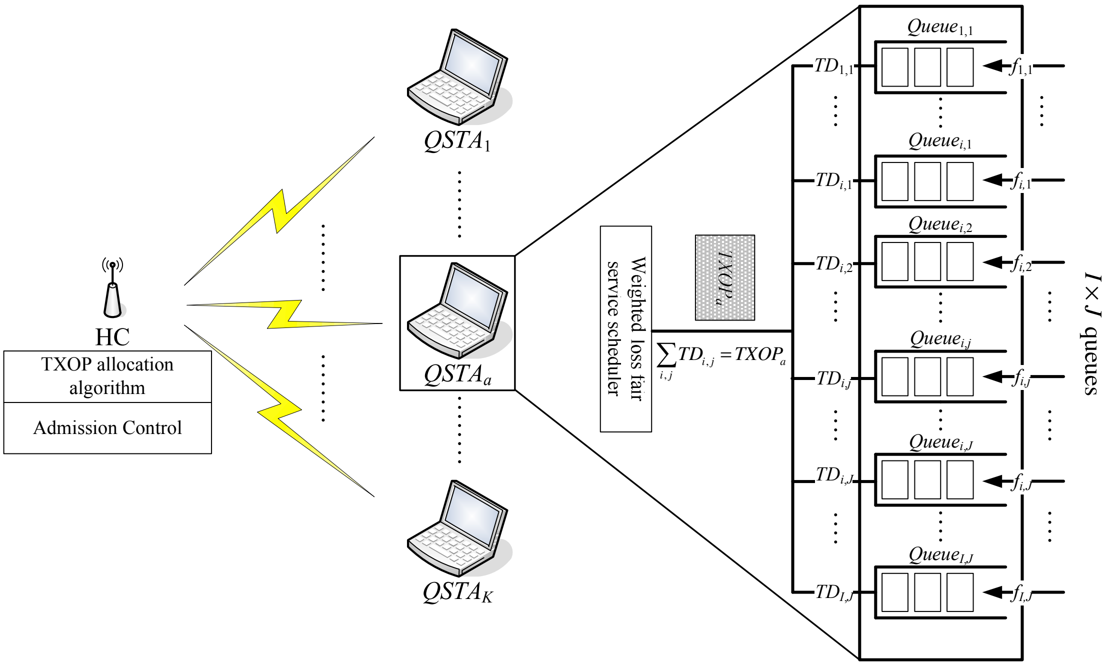

Fig. 2 The system architecture of our proposed scheme

Our proposed scheme consists of an aggregate TXOP allocation algorithm, the weighted-loss fair service scheduler, and the associated admission control unit. As mentioned before, TXOP allocation and admission control are performed in HC and weighted-loss fair service scheduler is implemented in QSTAs. An overview of our proposed scheme is depicted in Fig. 2. Once again, let us consider $QSTA_a$ with n_a traffic flows, which are classified into $I \times J$ groups according to their QoS requirements.

4.1 Aggregate TXOP Allocation Algorithm

First of all, an aggregate equivalent flow, denoted by $\hat{f}_i$, is determined using the technique described in the last section for flows $f_{i,1}$, $f_{i,2}$, …, and $f_{i,J}$, for all *i*,

$1 \le i \le I$. Note that the packet loss probability requirement of $\hat{f}_i$ is P_i. Let $N\left(\hat{\mu}_i, \hat{\sigma}_i^2\right)$ represent the traffic arrival distribution for flow $\hat{f}_i$. Define $\hat{f}$ as the ultimate equivalent flow with traffic arrival distribution $N\left(\sum_{i=1}^{I} \hat{\mu}_i, \sum_{i=1}^{I} \hat{\sigma}_i^2\right)$. The desired packet loss probability of flow $\hat{f}$, denoted by $P_{ultimate}$, is given by

$$P_{ultimate} = \frac{\sum_{i=1}^{I} P_i \cdot \hat{\mu}_i}{\sum_{i=1}^{I} \hat{\mu}_i} \tag{14}$$

Note that the delay bounds of the aggregate equivalent flows $\hat{f}_i$, $1 \le i \le I$, and the ultimate equivalent flow $\hat{f}$ are equal to SI. Consequently, the QoS parameter $\hat{\alpha}$ of flow $\hat{f}$ can be computed using equation (6) with desired packet loss probability $P_{ultimate}$. The aggregate TXOP allocated to $QSTA_a$ can be calculated using equation (10), except that the aggregate effective bandwidth and the average number of packets which can be served in one SI are obtained by

$$\hat{c} = \sum_{i=1}^{I} \hat{\mu}_i + \hat{\alpha} \sqrt{\sum_{i=1}^{I} \hat{\sigma}_i^2} \tag{15}$$

$$\overline{N} = \left\lceil \frac{\hat{c}}{\overline{L}} \right\rceil \tag{16}$$

In equation (16), $\overline{L}$ denotes the weighted average nominal packet size of all the flows in *F* and is calculated by $\overline{L} = \left(\sum_{i=1}^{I} \overline{N_i} \cdot \overline{L_i}\right) \Big/ \left(\sum_{i=1}^{I} \overline{N_i}\right)$, where $\overline{N_i}$ and $\overline{L_i}$ can be obtained using equations (12) and (13), respectively. The aggregate TXOP allocation procedure for $QSTA_a$ is summarized below.

Step 1. For $1 \le i \le I$, determine the aggregate equivalent flow $\hat{f}_i$ with packet loss probability requirement P_i for flows $f_{i,1}$, $f_{i,2}$, …, and $f_{i,J}$.

Step 2. Determine the packet loss probability $P_{ultimate}$ using equation (14).

Step 3. Compute the QoS parameter of the ultimate equivalent flow using equation

(6) with $P_{ultimate}$ as the desired packet loss probability.

Step 4. Compute the aggregate transmission duration $TXOP_a$ allocated to $QSTA_a$ using equation (10) with the effective bandwidth and average number of packets served in one SI obtained from equations (15) and (16), respectively.

4.2 Weighted-Loss Fair Service Scheduler

When polled, $QSTA_a$ needs to determine how the flows attached to it share the allocated TXOP. Let $Queue_{i,j}$ denote the queue maintained in $QSTA_a$ that is used to save packets of flow $f_{i,j}$. As shown in Fig. 3, $Queue_{i,j}$ is divided into β_j virtual sub-queues such that the p^{th} sub-queue, represented by $Queue_{i,j}^{p}$, $1 \le p \le \beta_j$, contains packets which can be kept for up to p SIs before violating the delay bound. How the allocated TXOP is shared is controlled by our proposed weighted-loss fair service scheduler.

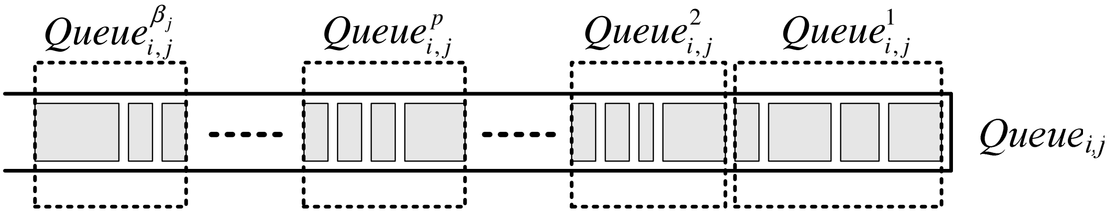

Fig. 3 The structure of sub-queues for $Queue_{i,j}$.

Consider the n^{th} SI. The proposed weighted-loss fair service scheduler is similar to the earliest deadline first (EDF) scheduler [17]. Let $Q_{i,j}^{p}[n]$, $1 \le p \le \beta_j$, represent the buffer occupancy in terms of transmission time for $Queue_{i,j}^{p}$ and $Q_{i,j}[n] = \sum_{p=1}^{\beta_j} Q_{i,j}^{p}[n]$. If the aggregate TXOP allocated to $QSTA_a$ satisfies $TXOP_a \ge \sum_{i,j} Q_{i,j}[n]$, then all packets in $Queue_{i,j}$ can be served and, therefore, no traffic is lost in the n^{th} SI. In this case, our proposed weighted-loss fair service scheduler is the same as the EDF scheduler.

Assume that $TXOP_a < \sum_{i,j} Q_{i,j}[n]$. Under this assumption, there exists a minimum m such that $\sum_{i,j}\sum_{p=1}^{m} Q_{i,j}^{p}[n] > TXOP_a$. Packets with deadlines smaller than $m \cdot SI$ are served in this SI according to the EDF scheduler. Any packet which can be kept for longer than $m \cdot SI$ stays in queue. Packets in $Queue_{i,j}^{m}$, $1 \le i \le I$, $\beta_j \ge m$, are handled differently by our proposed weighted-loss fair service scheduler and the EDF scheduler. In the proposed weighted-loss fair service scheduler, which packets should stay in queue (if $m > 1$) or be dropped (if $m = 1$) is decided based on running packet loss probabilities. Once the decision is made, the service order of those packets to be transmitted is determined by the EDF scheduler.

Define $Loss[n] = \sum_{i,j}\sum_{p=1}^{m} Q_{i,j}^{p}[n] - TXOP_a$. For $Queue_{i,j}$, let $A_{i,j}[n]$ and $L_{i,j}[n]$ denote, respectively, the accumulated amount of traffic arrived and lost up to the n^{th} SI. Define $l_{i,j}[n]$ as the amount of lost traffic (if $m = 1$) or the amount of traffic with deadline $m \cdot SI$ that stays in $Queue_{i,j}$ (if $m > 1$). Also, define $TD_{i,j}[n]$ as the TXOP duration shared by $Queue_{i,j}$. It holds that $\sum_{i,j} TD_{i,j}[n] = TXOP_a$. Finally, let $P_{i,j}[n] = \left(L_{i,j}[n-1] + l_{i,j}[n]\right) / A_{i,j}[n]$. We call $P_{i,j}[n]$ the running packet loss probability for $Queue_{i,j}$ up to the n^{th} SI if $m = 1$, or a pseudo one if $m > 1$.

Our proposed weighted-loss fair service scheduler tries to minimize the total amount of packet loss while maintaining a kind of fairness in the sense that the (pseudo) running packet loss probabilities of traffic flows are proportional to their packet loss probability requirements. To achieve the goal, we let $l_{i,j}[n]=0$ if $\beta_j < m$ or $\beta_j \geq m$ and $Q^m_{i,j}[n]=0$. For $Queue_{i,j}$ with $\beta_j \geq m$ and $Q^m_{i,j}[n]>0$, the following equations are solved for $l_{i,j}[n]$.

$$\frac{P_{i,j}[n]}{P_i}=\frac{P_{r,s}[n]}{P_r} \quad \forall (i,j),(r,s)\in U_{active} \tag{17}$$

$$Loss[n]=\sum_{(i,j)\in U_{active}} l_{i,j}[n] \tag{18}$$

In equations (17) and (18), U_{active} is a set which contains (i,j) such that $Q^m_{i,j}[n]>0$. For ease of description, we assume that every $Queue_{i,j}$ is in U_{active} if $\beta_j \geq m$. After some derivations (shown in the Appendix A), we get

$$\begin{aligned} l_{i,j}[n]=\frac{1}{\sum\limits_{(r,s)\in U_{active}} P_r\cdot A_{r,s}[n]}\Bigg[& P_i\cdot A_{i,j}[n]\cdot\Bigg(Loss[n]+\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}} L_{r,s}[n-1]\Bigg) \\ & -L_{i,j}[n-1]\cdot\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}} P_r\cdot A_{r,s}[n]\Bigg] \end{aligned} \tag{19}$$

If the solution satisfies $0\leq l_{i,j}[n]\leq Q^m_{i,j}[n]$ for all $(i,j)\in U_{active}$, then a feasible solution is obtained. The TXOP duration for $Queue_{i,j}$, i.e., $TD_{i,j}[n]$, is given by

$$TD_{i,j}[n]=\left(\sum_{p=1}^{m-1} Q^p_{i,j}[n]\right)+Q^m_{i,j}[n]-l_{i,j}[n] \tag{20}$$

However, the solution obtained by equation (19) may be infeasible, i.e., it is possible to have $l_{i,j}[n]>Q^m_{i,j}[n]$ or $l_{i,j}[n]<0$ for some $(i,j)\in U_{active}$. If it happens, then adjustment is necessary to make the solution feasible. The adjustment is accomplished by the loss computation algorithm shown in Appendix B. Its basic idea is described below. There are four possible cases for the solution obtained by equation (19).

<u>Case 1</u> $0 \le l_{i,j}[n] \le Q_{i,j}^{m}[n]$ for all $(i,j) \in U_{active}$.

If $0 \le l_{i,j}[n] \le Q_{i,j}^{m}[n]$ for all $(i,j) \in U_{active}$, then a feasible solution is found.

<u>Case 2</u> $l_{i,j}[n] \ge 0$ for all $(i,j) \in U_{active}$ and $l_{r,s}[n] > Q_{r,s}^{m}[n]$ for some (r,s).

In this case, let $Loss'[n] = Loss[n]$. For every (i,j) such that $l_{i,j}[n] \ge Q_{i,j}^{m}[n]$, assign $l_{i,j}[n] = Q_{i,j}^{m}[n]$, remove (i,j) from U_{active}, and set $Loss'[n] = Loss'[n] - Q_{i,j}^{m}[n]$. Use equation (19) again to compute $l_{i,j}[n]$ for the updated U_{active} and $Loss'[n]$. Note that, as proved in Theorem 1 below, the updated solution should fall in either Case 1 or Case 2. If it falls in Case 1, then a feasible solution is obtained. Otherwise, the same process is repeated. Eventually, a feasible solution will be obtained because it holds that $\sum_{i,j} Q_{i,j}^{m}[n] > Loss[n]$.

<u>Theorem 1</u> Given U_{active} and $Loss[n]$. Assume that the solution shown in equation (19) satisfies $l_{i,j}[n] \ge 0$ for all $(i,j) \in U_{active}$ and $l_{r,s}[n] > Q_{r,s}^{m}[n]$ for some (r,s). Let $U = U_{active} - \{(r,s)\}$ and $Loss'[n] = Loss[n] - Q_{r,s}^{m}[n]$. Further, let $l'_{i,j}[n]$, $(i,j) \in U$, be the solution of equations (17) and (18) for U and $Loss'[n]$. It holds that $l'_{i,j}[n] > l_{i,j}[n] > 0$.

Proof: Assume that $l'_{i,j}[n] \le l_{i,j}[n]$ for some (i,j). According to equation (17), we have $l'_{a,b}[n] \le l_{a,b}[n]$ for any $(a,b) \in U$. As a result, it holds that $\sum_{(a,b)\in U} l'_{a,b}[n] \le \sum_{(a,b)\in U} l_{a,b}[n] = \sum_{(a,b)\in U_{active}} l_{a,b}[n] - l_{r,s}[n] < \sum_{(a,b)\in U_{active}} l_{a,b}[n] - Q_{r,s}^{m}[n]$ $= Loss'[n]$. This contradicts equation (18). Therefore, Theorem 1 is true.

Theorem 1 says that if we set $l_{r,s}[n] = Q_{r,s}^{m}[n]$ when $l_{r,s}[n] > Q_{r,s}^{m}[n]$, then $l_{i,j}[n]$ has to be increased for all $(i,j) \in U$ in order to satisfy equation (17) for queues in U and equation (18). In fact, the amount $l_{r,s}[n]-Q^m_{r,s}[n]$ is proportionally shared by queues in U, i.e., it holds that $(l'_{a,b}[n]-l_{a,b}[n])/A_{a,b}[n]P_a=(l'_{c,d}[n]-l_{c,d}[n])/A_{c,d}[n]P_c$ for all $(a,b),\ (c,d)\in U$. It is worth to point out that although Theorem 1 is stated for one (r,s) which satisfies $l_{r,s}[n]>Q^m_{r,s}[n]$, it actually implies the same conclusion if multiple queues satisfy the condition.

<u>Case 3</u> $l_{i,j}[n]<Q^m_{i,j}[n]$ for all $(i,j)\in U_{active}$ and $l_{r,s}[n]<0$ for some (r,s).

In this case, we assign $l_{i,j}[n]=0$ for every (i,j) such that $l_{i,j}[n]\le 0$, remove (i,j) from U_{active}, and solve for new $l_{i,j}[n]$ with equation (19) for the updated U_{active} and $Loss[n]$. The updated solution will fall in either Case 1 or Case 3. This is implied by Theorem 2 stated below. Similarly, a feasible solution is found if the updated solution falls in Case 1. Otherwise, the same process is repeated till a feasible solution appears. The proof for Theorem 2 is similar to that for Theorem 1 and is omitted.

<u>Theorem 2</u> Given U_{active} and $Loss[n]$. Assume that the solution shown in equation (19) satisfies $l_{i,j}[n]\le Q^m_{i,j}[n]$ for all $(i,j)\in U_{active}$ and $l_{r,s}[n]<0$ for some (r,s). Let $U=U_{active}-\{(r,s)\}$ and $l'_{i,j}[n]$, $(i,j)\in U$, be the solution of equations (17) and (18) for U and $Loss[n]$. It holds that $l'_{i,j}[n]<l_{i,j}[n]<Q^m_{i,j}[n]$.

Theorem 2 states that if we set $l_{r,s}[n]=0$ when $l_{r,s}[n]<0$, then $l_{i,j}[n]$ has to be decreased for all $(i,j)\in U$ in order to satisfy equation (17) for queues in U and equation (18). Again, although we state Theorem 2 for one (r,s) which satisfies

$l_{r,s}[n] < 0$, it implies the same conclusion if multiple queues satisfy the condition. Therefore, for Case 3, we can repeatedly set $l_{i,j}[n] = 0$ for all (i, j) such that $l_{i,j}[n] \le 0$ and solve equations (17) and (18) for the updated U_{active} and $Loss[n]$ until a feasible solution is found.

<u>**Case 4**</u> $l_{r,s}[n] > Q_{r,s}^{m}[n]$ for some (r, s) and $l_{r',s'}[n] < 0$ for some (r', s').

Let U be the set which contains all $(i, j) \in U_{active}$, such that $l_{i,j}[n] \ge 0$. Case 4 is further divided into two sub-cases.

<u>**Sub-case 1**</u> $\sum_{(i,j)\in U} Q_{i,j}^{m}[n] < Loss[n]$

For this sub-case, define $V_1 = \{(i, j) \in U_{active} : l_{i,j}[n] \ge Q_{i,j}^{m}[n]\}$ and $V_2 = U_{active} - V_1$. We set $l_{i,j}[n] = Q_{i,j}^{m}[n]$ for all $(i, j) \in V_1$ and $Loss'[n] = Loss[n] - \sum_{(i,j)\in V_1} l_{i,j}[n]$. Then, solve equations (17) and (18) for V_2 and $Loss'[n]$. Let $l'_{i,j}[n]$, $(i, j) \in V_2$, be the solution. No further adjustment is necessary if the solution falls in Case 1. If the solution falls in Case 2, then Case 2 is performed repeatedly until a feasible solution is found. Similarly, if the solution falls in Case 3, then Case 3 will be repeatedly executed until a feasible solution is obtained. Finally, if the solution falls in Case 4, then either Sub-case 1 or Sub-case 2 is performed again.

<u>**Sub-case 2**</u> $\sum_{(i,j)\in U} Q_{i,j}^{m}[n] \ge Loss[n]$

For this sub-case, let $V_1 = U$ and $V_2 = U_{active} - V_1$. Equations (17) and (18) are solved for V_1 and $Loss[n]$. If the solution falls in Case 1, then no further processing is required. Assume that the solution falls in Case 2. Let $l'_{i,j}[n]$,

$(i, j) \in V_1$, be the solutions and W_1 and W_2 be two sub-sets of V_1 such that $W_1 = \{(i, j) \in V_1 : l'_{i,j}[n] < Q^m_{i,j}[n]\}$ and $W_2 = V_1 - W_1$. We set $l_{i,j}[n] = Q^m_{i,j}[n]$ for all $(i, j) \in W_2$. Let $V_2 = V_2 \cup W_1$ and $Loss'[n] = Loss[n] - \sum_{(i,j)\in W_2} l_{i,j}[n]$. Equations (17) and (18) are solved for V_2 and $Loss'[n]$. Note that this step is necessary to achieve the equality described in equation (17) for queues in the updated V_2. If the solution falls in Case 3, then Case 3 will be repeatedly executed until a feasible solution is obtained. Finally, if the solution falls in Case 4, then either Sub-case 1 or Sub-case 2 is performed again.

The computational complexity of the loss computation algorithm is stated in the following Theorem 3.

Theorem 3 The loss computation algorithm takes at most $2(N-1)$ iterations to find the feasible solution, where $N = |U_{active}|$, the size of U_{active}.

Proof: It is clear that the solution of the last iteration falls in Case 1. Let *M* denote the size of *U* in that iteration. We shall prove that the loss computation algorithm takes at most $2(N-M)$ iterations to find the feasible solution if $M < N$ or one iteration if $M = N$. The case of $M = N$ is obviously true. We prove the case of $M < N$ by mathematical induction. For simplicity, we use Sub-case *i* ($i = 1, 2$) to represent Sub-case *i* of Case 4 in this proof.

For $N = 2$, we have $M = 1$. Since $M < N$, we know that the solution of the first iteration cannot fall in Case 1. By tracing the algorithm, one can see that the number of iterations required to find the feasible solution is equal to $2 = 2(N-M)$.

Assume that the statement is true for $N = H$ and $M = 1, 2, ..., H-1$ (Hypothesis I). Consider the case of $N = H+1$. If Sub-case 2 is never visited, then the number of iterations required is at most $N - M + 1 \leq 2(N - M)$ because at least one queue is removed from U_{active} in each iteration before the last one. Assume that Sub-case 2 was visited before the feasible solution is found. If the solution of the first iteration does not fall in Sub-case 2, then the size of *U* in the second iteration is at most *H*. According to Hypothesis I, the maximum number of iterations required to find the feasible solution, starting from iteration 2, is equal to $2(H - M)$. As a result, the total number of iterations is upper bounded by $2(H - M) + 1 < 2(N - M)$.

Assume that the solution of the first iteration falls in Sub-case 2. Let $|V_1| = i$ and $|V_2| = j$ with $i + j = N$. Further, let k represent the number of queues added to V_2 when iteration 1 resumes its execution. The total number of iterations required is at most $1 + B(i,k) + 2(j + k - M)$, where $B(i,k)$ represents the maximum number of iterations required before iteration 1 resumes its execution and $2(j + k - M)$ denotes the upper bound of the number of iterations required to find the feasible solution for the updated V_2, according to Hypothesis I. Theorem 3 is true if $B(i,k) \leq 2(i - k) - 1$. We shall prove this by mathematical induction.

By tracing the algorithm one can see that it is true for $i = 2$ and $k = 0$ or 1. Assume that it is true for $i = p$ and $k = 0, 1, ..., p-1$ (Hypothesis II). Consider the case of $i = p+1$. If Sub-case 2 is not visited again before iteration 1 resumes its execution, then we have $B(i,k) \leq i - k$. Note that if $k = 0$, then Case 2 is not visited. If $k > 0$, then there are 0 to $(i - k - 1)$ times of Sub-case 1 followed by a

Case 2. Since $k \le i-1$, we have $B(i,k) \le 2(i-k)-1$. Assume that, before Sub-case 1 resumes its execution, Sub-case 2 is visited for the second time in iteration *r*. This implies the solutions of iterations 2, …, and $r-1$ all fall in Sub-case 1 and, therefore, at least $r-2$ queues are removed from U_{active}. Let x, *y*, and *z* represent, respectively, the size of V_1, the size of V_2, and the number of queues added to V_2 when iteration *r* resumes its execution. It is clear that $x+y \le i-r+2$. After iteration *r* resumes its execution, the situation is the same as iteration 1 except that the size of V_1 (of iteration 1) is changed from i to $y+z$. As a result, we have $B(i,k) \le (r-1)+B(x,z)+B(y+z,k)$. According to Hypothesis II, it holds that $B(i,k) \le \ (r-1)+2(x-z)-1+2(y+z-k)-1 \le 2(i-k)-1$. This completes the proof of Theorem 3.

After the feasible solution is found, $TD_{i,j}[n]$ can be obtained according to equation (20). If data are dropped (i.e., $m=1$), $L_{i,j}[n]$ is updated as follows

$$L_{i,j}[n] = L_{i,j}[n-1] + l_{i,j}[n] \tag{21}$$

Since the number of real-time flows attached to each QSTA is normally small, the complexity of the loss computation algorithm should be acceptable. Furthermore, because of static and periodic TXOP allocation, each QSTA has time one *SI* to compute the solution. Therefore, the proposed weighted-loss fair service scheduler should be feasible for real systems.

4.3 The Associated Admission Control Unit

Assume that *QSTA*$_a$ is negotiating with HC for a new traffic flow, i.e., the

$(n_a+1)^{th}$ flow of *QSTA*$_a$, that requires packet loss probability P_i and delay bound D_j. Define available bandwidth BW_{ava} as

$$BW_{ava} = SI\left(1-\frac{T_{cp}}{T_b}\right)-\sum_{i=1}^{K} TXOP_i \tag{22}$$

Let θ and ρ^2 denote, respectively, the mean and variance of traffic arrival in one SI for the new traffic flow. The new flow, if admitted, will become part of flow $f_{i,j}$. As a result, we need only update the parameters related to flows $f_{i,j}$, $\hat{f}_i$ and $\hat{f}$. Let $N\left(\mu'_{i,j},\sigma'^2_{i,j}\right)$, $N\left(\hat{\mu}'_{i,j},\hat{\sigma}'^2_{i,j}\right)$ and $N\left(\hat{\mu}'_i,\hat{\sigma}'^2_i\right)$ denote, respectively, the traffic arrival distributions for flows $f_{i,j}$, $\hat{f}_{i,j}$ and $\hat{f}_i$ before the new flow is admitted. Assume that this new flow is admitted. The parameters of $f_{i,j}$ are updated as $\mu_{i,j} = \mu'_{i,j}+\theta$ and $\sigma^2_{i,j} = \sigma'^2_{i,j}+\rho^2$. Moreover, the traffic arrival distribution of the aggregate equivalent flow $\hat{f}_i$ is updated as $N\left(\hat{\mu}_i,\hat{\sigma}^2_i\right)$, where $\hat{\mu}_i = \mu_{i,j}+\sum_{s\neq j,s=1}^{J}\hat{\mu}'_{i,s}$ and $\hat{\sigma}^2_i = \left(\alpha_{i,j}\sigma_{i,j}/\hat{\alpha}_{i,j}\right)^2+\sigma'^2_{i,1}+\sum_{s\neq j,s=2}^{J}\hat{\sigma}'^2_{i,s}$ (if $j\neq 1$) or $\hat{\sigma}^2_i = \sigma^2_{i,1}+\sum_{s=2}^{J}\hat{\sigma}'^2_{i,s}$ (if $j=1$). The traffic arrival distribution of the ultimate equivalent flow $\hat{f}$ is updated as $N\left(\hat{\mu},\hat{\sigma}^2\right)$, where $\hat{\mu} = \hat{\mu}_i+\sum_{r\neq i,r=1}^{I}\hat{\mu}'_r$ and $\hat{\sigma}^2 = \hat{\sigma}^2_i+\sum_{r\neq i,r=1}^{I}\hat{\sigma}'^2_r$. The ultimate packet loss probability has to be recalculated using equation (14) with the above updated parameters as input. Finally, the effective bandwidth and the required TXOP, denoted by $TXOP^*_a$, can be computed, respectively, by equations (6) and (10). Define $\Delta TXOP = TXOP^*_a - TXOP_a$. The new flow is admitted iff the following inequality is satisfied.

$$BW_{ava} \geq \Delta TXOP \tag{23}$$

If the new flow is admitted, we update BW_{ava} by $BW_{ava} = BW_{ava} - \Delta TXOP$.

Note that, if an existing flow of $QSTA_a$ is disconnected, a process similar to that shown above is conducted to obtain $\Delta TXOP = TXOP_a - TXOP_a^*$, and BW_{ava} is updated by $BW_{ava} = BW_{ava} + \Delta TXOP$. Note that if admission or disconnection of a flow leads to change of SI, then the TXOPs for all QSTAs should be recalculated.

5. Simulation Results

The PHY and MAC parameters and all related information used in simulations are shown in Table I. Note that the sizes of QoS-ACK and QoS-Poll in the table only include the sizes of MAC header and CRC overhead.

Table I Related parameters used in simulations.

PHY and MAC parameters	
SIFS	10 us
MAC Header size	32 bytes
CRC size	4 bytes
QoS-ACK frame size	16 bytes
QoS CF-Poll frame size	36 bytes
PLCP Header Length	4 bytes
PLCP Preamble length	20 bytes
PHY rate(R)	11 Mbps
Minimum PHY rate (R_{min})	2 Mbps
Transmission time for different header and per-packet overhead	
PLCP Preamble and Header (t_{PLCP})	96 μs
Data MAC Header (t_{HDR})	23.2727 μs
Data CRC (t_{CRC})	2.90909 μs
ACK frame (t_{ACK})	107.63636 μs
QoS-CFPoll (t_{POLL})	122.1818 μs
Per-packet overhead (O)	249.81818 μs

Traffic is delivered from QSTAs to AP and the contention-free period occupies the whole SI. We investigate three types of QSTA in the simulation. Each type of QSTA is assumed to be attached with two real-time traffic flows. Real traffic traces [18] are used in our simulations. The length of each real traffic trace lasts for one hour. The detailed information of traffic flows, including QoS requirements and traffic parameters, are described in Table II. For each flow, the mean μ and the variance σ^2 of traffic arrivals in one SI can be calculated from the mean data rate

ρ and the variance of frame size υ^2 provided in the trace file, using the technique described in Section 2. The calculated μ and σ^2 of each flow are shown in the last two rows of Table II. Because of the static and periodic nature of our proposed scheme, the performance of each type of QSTA is independent of whether or not there exists any other QSTA in the system. Therefore, to evaluate the efficiency of the proposed TXOP allocation algorithm, it suffices to study a system which consists of one HC and three QSTAs, one for each type. Note that Type III QSTA is included to study the effect of aggregating flows with identical QoS requirements.

Table II TSPECs of traffic flows attached to Type I, Type II and Type III QSTAs.

Type of QSTA	Type I		Type II		Type III	
Attached Real Trace	Jurassic Park I	Lecture Camera	Mr. Bean	Office Camera	Lecture Camera	Office Camera
Packet Loss Rate Requirement (P_L)	0.01	0.001	0.01	0.001	0.001	0.001
Maximum Service Interval (SI_{max})	80(ms)	160(ms)	80(ms)	160(ms)	160(ms)	160(ms)
Mean Data Rate (ρ)	268k(bps)	210k(bps)	184k(bps)	112k(bps)	210k(bps)	112k(bps)
Nominal MSDU size (L)	1339 (bytes)	1048 (bytes)	920(bytes)	558(bytes)	1048 (bytes)	558(bytes)
Variance of Frame Size (υ^2)	1273237	828990	801216	1604797	828990	1604797
Frame inter-arrival time	40 (ms)					
Scheduled Service Interval (SI)	80 (ms)					
Calculated Mean per SI (μ)	2680 (bytes)	2100 (bytes)	1840 (bytes)	1120 (bytes)	2100 (bytes)	1120 (bytes)
Calculated Variance per SI (σ^2)	2546474	1657980	1602432	3209594	1657980	3209594

Table III The 99% confidence intervals of packet loss probability of flows attached to Type I, Type II and Type III QSTAs.

	Packet Loss Probability (P_L)					
	Type I QSTA		Type II QSTA		Type III QSTA	
	Jurassic Park I	Lecture Camera	Mr. Bean	Office Camera	Lecture Camera	Office Camera
Sample scheduler	$0.1857 \pm 4\times10^{-3}$	$0.0186 \pm 3\times10^{-3}$	$0.2323 \pm 3\times10^{-5}$	$0.0232 \pm 6\times10^{-6}$	$0.0760 \pm 8\times10^{-5}$	$0.0760 \pm 9\times10^{-5}$
RVAC	$0.0008 \pm 4\times10^{-6}$	$0.0001 \pm 4\times10^{-6}$	$0.0025 \pm 9\times10^{-6}$	$0.0003 \pm 9\times10^{-6}$	$0.0002 \pm 8\times10^{-5}$	$0.0002 \pm 6\times10^{-5}$
Scheme of [7]	$0.0052 \pm 2\times10^{-5}$	$0.0005 \pm 2\times10^{-5}$	$0.0032 \pm 1\times10^{-5}$	$0.0003 \pm 1\times10^{-5}$	$0.0004 \pm 8\times10^{-5}$	$0.0004 \pm 1\times10^{-4}$
Our proposed scheme	$0.0099 \pm 3\times10^{-5}$	$0.0010 \pm 3\times10^{-5}$	$0.0072 \pm 2\times10^{-5}$	$0.0007 \pm 2\times10^{-5}$	$0.0004 \pm 8\times10^{-5}$	$0.0004 \pm 1\times10^{-4}$
Our proposed scheme*	$0.0100 \pm 1\times10^{-4}$	$0.0010 \pm 3\times10^{-5}$	$0.0073 \pm 1\times10^{-4}$	$0.0007 \pm 2\times10^{-5}$	$0.0004 \pm 8\times10^{-4}$	$0.0004 \pm 1\times10^{-3}$

The aggregate TXOP duration allocated by the sample scheduler is calculated by plugging the simulation parameters into equations (1) and (2). The aggregate TXOP duration for RVAC [8] is obtained assuming that all flows request the most stringent packet loss probability and delay bound because it only considers traffic flows with identical QoS requirements. For the scheme presented in [7], the aggregate TXOP allocation is calculated assuming all flows request the most stringent packet loss probability. For the scheme proposed in this project, the aggregate TXOP duration is derived according to the procedure described in Section 4. In the comparison, we adopt the proposed weighted-loss fair scheduler for all the investigated schemes and drop a packet iff it violates the delay bound.

In Table III, we compare packet loss probabilities after all data are delivered. Since there is only one trace for each video, we conducted simulations with 1,000 different starting positions to collect the 99% confidence intervals. The symbol $a \pm b$ in Table III means the 99% confidence interval is given by $(a-b, a+b)$. Transmission error is also considered for our proposed scheme. The frame transmission error probability is set to be 0.5×10^{-3}. The packet loss probability considering transmission error is marked with * and shown in the last row of Table III. According to the results, our proposed scheme can meet the individual QoS requirements requested by traffic flows whether or not there is aggregation of flows with identical QoS requirements. Moreover, no matter which TXOP allocation scheme is adopted, our proposed weighted-loss fair scheduler can achieve the goal of maintaining the ratio of actual packet loss probabilities as that of the requested values.

For example, the ratios of the actual packet loss probabilities of Jurassic Park I and Lecture Camera for the sample scheduler, the RVAC scheme, the scheme proposed in [7], our proposed scheme, and our proposed scheme with transmission error are, respectively, 0.1857:0.0186, 0.0008:0.0001, 0.0052:0.0005, 0.0099:0.0010, and 0.0100:0.0010, which are all very close to the ratio of the requested packet loss probabilities, i.e., 0.01:0.001. Another important observation is that the results of our proposed scheme are satisfactory even for a frame error rate of 0.5×10^{-3}. This implies that, to cope with transmission errors, one need only select an appropriate feasible physical transmission rate so that the probability of transmission error is sufficiently smaller than the requested packet loss probability.

We also record the running packet loss probabilities of traffic flows attached to Type I QSTA for all investigated schemes. Here, the running packet loss probability for flow $f_{i,j}$ up to the n^{th} SI is given by $L_{i,j}[n]/A_{i,j}[n]$. For sample scheduler, as shown in Fig. 4, the running packet loss probabilities of all simulated traffic flows are more than 10 times larger than their requested levels for most of the time. For TXOP allocation schemes which consider packet loss probability, we compare the sample paths of each traffic flow attached to Type I QSTA . The results are illustrated in Fig. 5 and Fig. 6. It can be seen that the long-term packet loss probability meets the requirement for all the investigated schemes. However, our proposed scheme is the most efficient one because it allocates the smallest TXOP durations to QSTAs. Fig. 7 compares the admissible regions of the investigated TXOP allocation schemes. For a particular scheme, the system can accommodate *x*

Type I QSTAs and *y* Type II QSTAs with QoS guarantee if (*x*, *y*) falls in the triangle formed by the *x*-axis, *y*-axis, and the curve labeled for the scheme. Our proposed scheme allows 8% and 18% more QSTAs to be admitted than the scheme proposed in [7] and RVAC, respectively.

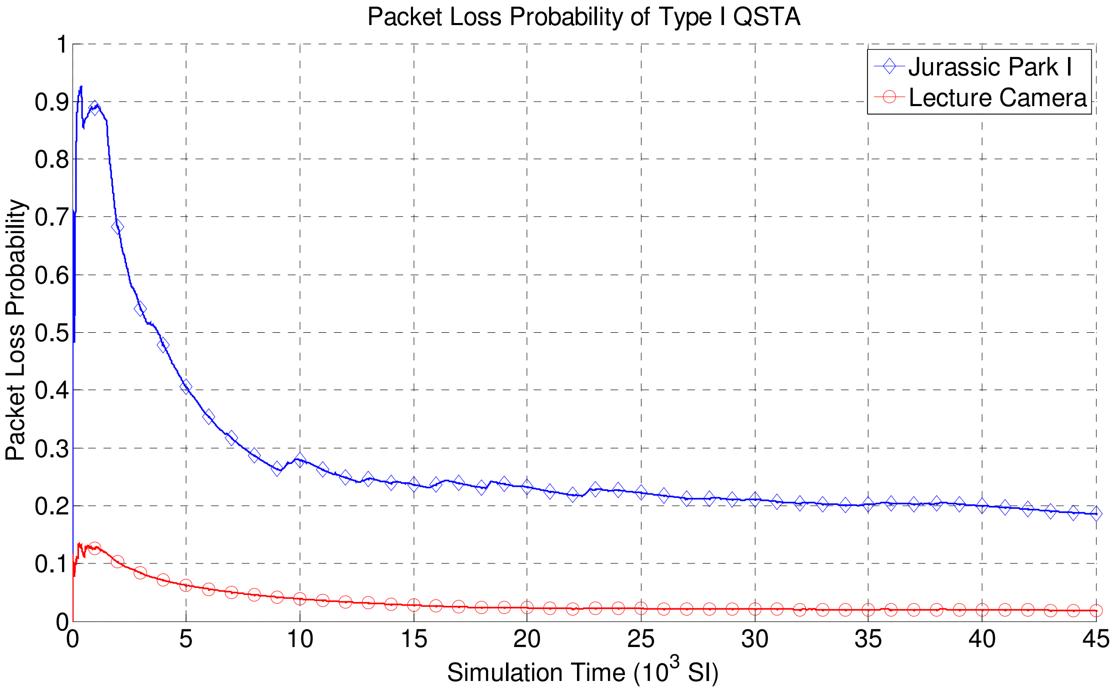

Fig. 4 Running packet loss probabilities of flows attached to Type I QSTA for the sample scheduler.

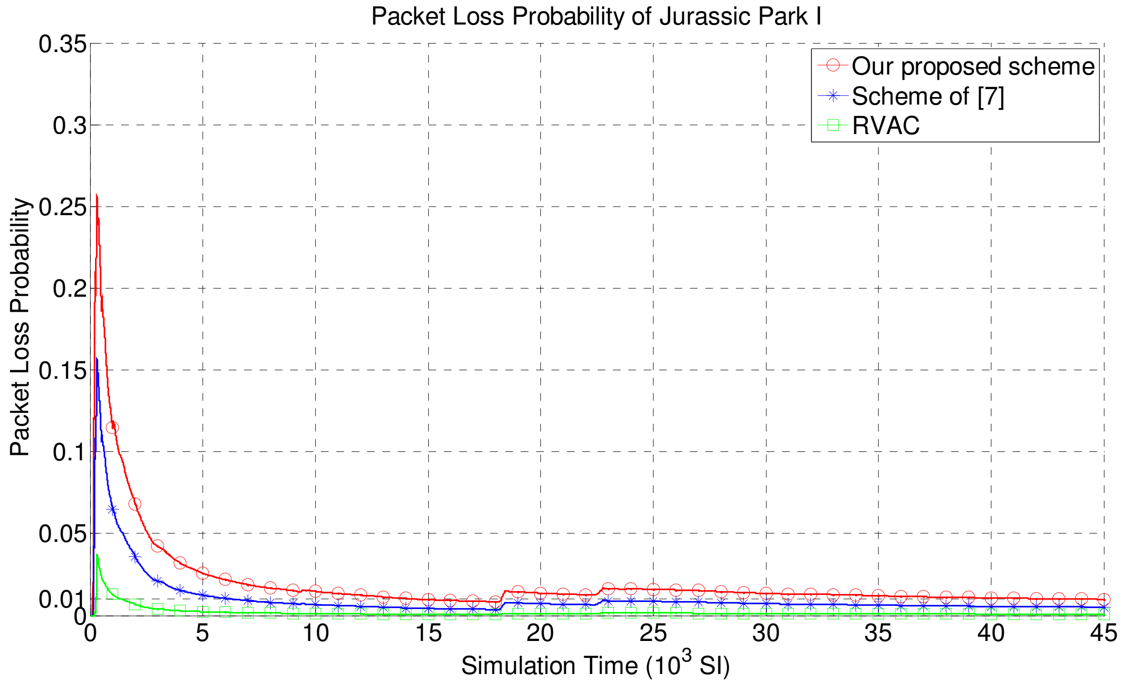

Fig. 5 Running packet loss probabilities of Jurassic Park I attached to Type I QSTA.

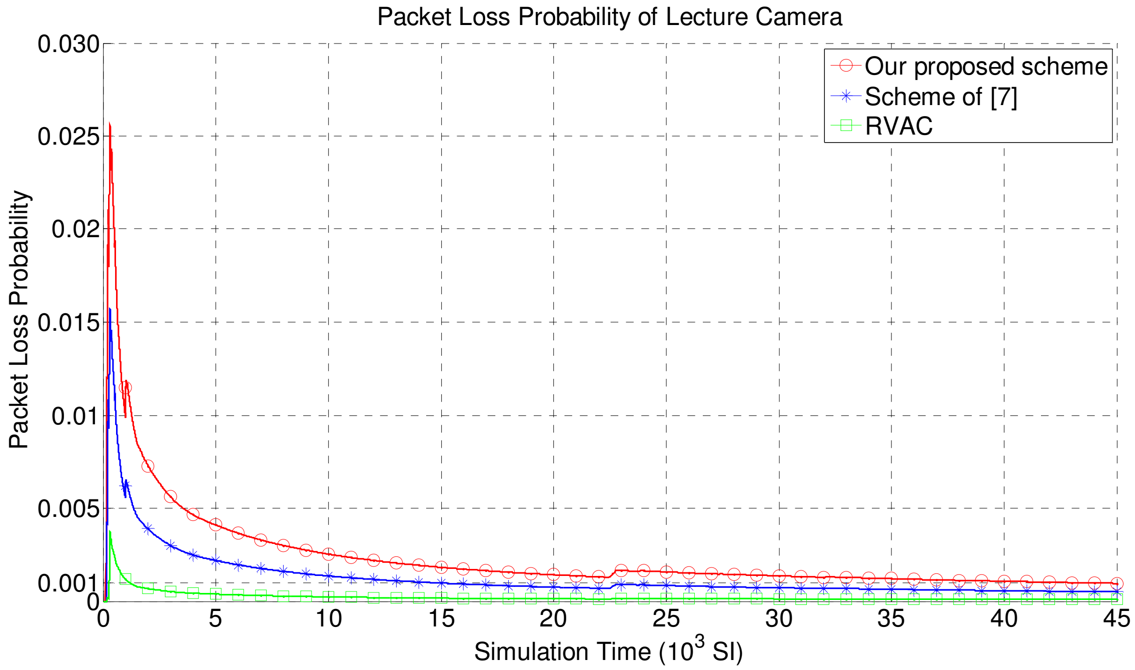

Fig. 6 Running packet loss probabilities of Lecture Camera attached to Type I QSTA.

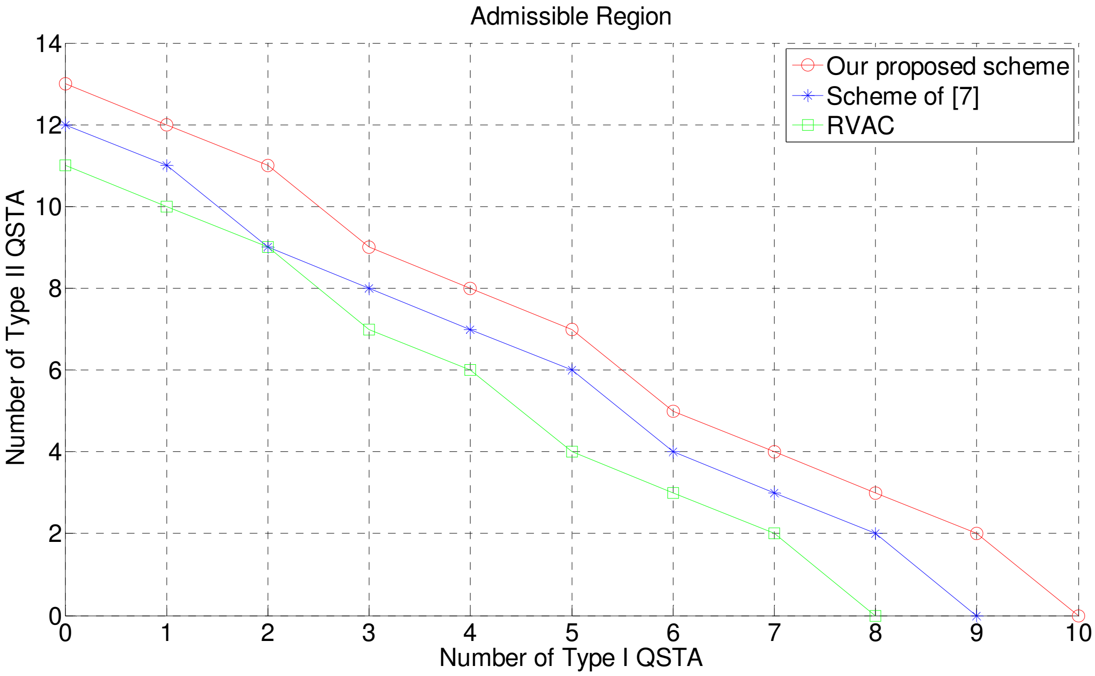

Fig. 7 Comparison of admissible regions.

6. Conclusion

In this project, we have presented an efficient TXOP allocation algorithm, a weighted-loss fair service scheduler, and the associated admission control unit to provide QoS guarantee for VBR traffic flows with different packet loss probability and delay bound requirements in WLANs. Computer simulations are conducted to evaluate the performance of our proposed scheme. Results show that our proposed scheme is effective in QoS guarantee and, moreover, performs much better than previous works. In real systems, it is likely that there are only a limited number of possible applications. Therefore, one can pre-compute the QoS parameter of each type of application so that admission control can be performed in real time. An interesting further research topic is to mathematically analyze the performance of the proposed scheme for various traffic models and QoS requirements.

7. Appendix

Derivation of equation (19)

As defined in Section 4, the running packet loss probability of $f_{i,j}$, namely, $P_{i,j}[n]$, can be written as

$$P_{i,j}[n]=\frac{L_{i,j}[n-1]+l_{i,j}[n]}{P_i\cdot A_{i,j}[n]}$$

After substituting the above equation into equation (17), we get

$$\frac{L_{i,j}[n-1]+l_{i,j}[n]}{P_i\cdot A_{i,j}[n]}=\frac{L_{r,s}[n-1]+l_{r,s}[n]}{P_r\cdot A_{r,s}[n]},$$

which implies

$$l_{r,s}[n]=-L_{r,s}[n-1]+\left(\frac{P_r\cdot A_{r,s}[n]}{P_i\cdot A_{i,j}[n]}\right)\left(L_{i,j}[n-1]+l_{i,j}[n]\right).$$

Summing over all $(r,s)\in U_{active}$ except for $(r,s)=(i,j)$, we have

$$\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}} l_{r,s}[n]=\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}}\left[-L_{r,s}[n-1]+\left(\frac{P_r\cdot A_{r,s}[n]}{P_i\cdot A_{i,j}[n]}\right)\left(L_{i,j}[n-1]+l_{i,j}[n]\right)\right]$$

.

According to equation (18), it holds that

$$Loss[n]-l_{i,j}[n]=\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}}\left[-L_{r,s}[n-1]+\left(\frac{P_r\cdot A_{r,s}[n]}{P_i\cdot A_{i,j}[n]}\right)\left(L_{i,j}[n-1]+l_{i,j}[n]\right)\right].$$

After some manipulations, we get

$$l_{i,j}[n]=\frac{1}{\sum_{(r,s)\in U_{active}} P_r\cdot A_{r,s}[n]}\cdot\left[P_i\cdot A_{i,j}[n]\cdot\left(Loss[n]+\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}} L_{r,s}[n-1]\right)\right.$$

$$\left.-L_{i,j}[n-1]\cdot\left(\sum_{(r,s)\neq(i,j),(r,s)\in U_{active}} P_r\cdot A_{r,s}[n]\right)\right].$$

Pseudo code for computing feasible $l_{i,j}[n]$ for all $(i,j) \in U_{active}$.

```
Algorithm: Loss computation
Initialization
$U_{temp} = U_{active}$
$Loss_{temp} = Loss[n]$
$Flag = 0$
Begin
$\left[l_{i,j}[n], \forall(i,j) \in U_{active}\right]_{1\times|U_{active}|} = LossComputation(Loss_{temp}, U_{temp})$
End
/*Loss computation module*/
$LossComputation(Loss, U)$
    $WeightedLossCalculation(Loss, U)$ /*Compute $l_{i,j}[n]$ with eqn. (19)*/
    if $0 \le l_{i,j}[n] \le Q^m_{i,j}[n]$ $\forall(i,j) \in U$ /*Case 1*/
        exit
    elseif $0 \le l_{i,j}[n]$ $\forall(i,j) \in U$ and $\exists(i,j) \in U, s.t.\ l_{i,j}[n] > Q^m_{i,j}[n]$ /*Case 2*/
        for all $(i,j) \in U$
            if $l_{i,j}[n] \ge Q^m_{i,j}[n]$
                $l_{i,j}[n] = Q^m_{i,j}[n]$
                $U = U - \{(i,j)\}$
                $Loss = Loss - l_{i,j}[n]$
            end if
        end for
        if $Flag = 1$
            $Flag = 0$
            exit
        else
            $LossComputation(Loss, U)$
        end if
    elseif $l_{i,j}[n] \le Q^m_{i,j}[n]$ $\forall(i,j) \in U$ and $\exists(i,j) \in U, s.t.\ l_{i,j}[n] < 0$ /*Case 3*/
        for all $(i,j) \in U$
            if $l_{i,j}[n] \le 0$
                $l_{i,j}[n] = 0$
                $U = U - \{(i,j)\}$
            end if
        end for
        $LossComputation(Loss, U)$
    else /*Case 4: $\exists(i,j)$ and $(r,s) \in U$, $s.t.\ l_{i,j}[n] > Q^m_{i,j}[n]$ and $l_{r,s}[n] < 0$ */
        $V_1 = \{(i,j) \in U: l_{i,j}[n] \ge 0\}$
        $V_2 = U - V_1$
        **if** $\sum_{(i,j)\in V_1} Q_{i,j}^m[n] < Loss[n]$  /*Sub-case 1*/
            **for all** $(i,j)\in U$
                **if** $l_{i,j}[n] \geq Q_{i,j}^m[n]$
                    $l_{i,j}[n] = Q_{i,j}^m[n]$
                    $U = U - \{(i,j)\}$
                    $Loss = Loss - l_{i,j}[n]$
                **end if**
            **end for**
            $LossComputation(Loss, U)$
        **else** /*Sub-case 2: $\sum_{(i,j)\in V_1} Q_{i,j}^m[n] \geq Loss[n]$ */
            $Flag = 1$
            $LossComputation(Loss, V_1)$
            **if** $Flag = 0$ and $\exists(i,j)\in V_1, s.t.\ l_{i,j}[n] < Q_{i,j}^m[n]$
                **for all** $(i,j)\in V_1$
                    **if** $l_{i,j}[n] < Q_{i,j}^m[n]$
                        $V_2 = V_2 \cup \{(i,j)\}$
                    **else**
                        $Loss = Loss - Q_{i,j}^m[n]$
                    **end if**
                **end for**
                $LossComputation(Loss, V_2)$
            **else**
                **for all** $(i,j)\in V_2$
                    $l_{i,j}[n] = 0$
                **end for**
                **exit**
            **end if**
        **end if**
    **end if**
```

Summary of notation used in this project

	The investigated QoS requirements	
1. P_i	The i^{th} packet loss probability requirement.	$1 \le i \le I$
2. D_j	The j^{th} delay bound requirement ($D_j = \beta_j \cdot SI$).	$1 \le j \le J$
	Traffic flow notation identified by QoS requirements	
3. $F_{i,j}$	The set which contains traffic flows with packet loss probability and delay bound requirement equal to P_i and D_j ,respectively.	$1 \le i \le I$, $1 \le j \le J$
4. $f_{i,j}$	The aggregate traffic flow for all flows in $F_{i,j}$ with distribution assumed to be $N(\mu_{i,j}, \sigma_{i,j}^2)$.	
5. F_i	The set which contains traffic flows with packet loss probability equal to P_i (i.e., $F_i = \bigcup_{j=1}^{J} F_{i,j}$).	$1 \le i \le I$
6. f_i	The aggregate traffic flow for all flows in F_i .	
7. F	The set which all traffic flows.	
8. f	The aggregate traffic flow for all flows in *F*.	
	Traffic flow specification for $f_{i,j}$	
9. $\alpha_{i,j}$	QoS parameter of $f_{i,j}$.	$1 \le i \le I$, $1 \le j \le J$
10. $c_{i,j}$	Effective bandwidth of $f_{i,j}$ in one SI. (i.e., $c_{i,j} = \mu_{i,j} + \alpha_{i,j}\sigma_{i,j}$)	
11. $N_{i,j}$	The number of packets belonging to $f_{i,j}$ arrived in one SI.	
12. $\overline{N_{i,j}}$	The estimated number of packets belonging to $f_{i,j}$ arrived in one SI.	
13. $X_{i,j}$	Packet size distribution of $f_{i,j}$.	
14. $L_{i,j}$	Nominal MSDU size of $f_{i,j}$.	
	Equivalent traffic flow notation identified by QoS requirements	
15. $\hat{f}_{i,j}$	The equivalent flow of $f_{i,j}$ with distribution denoted by $N(\hat{\mu}_{i,j}, \hat{\sigma}_{i,j}^2)$.	$1 \le i \le I$, $1 \le j \le J$
16. $\hat{f}_i$	The equivalent flow of f_i with distribution denoted by $N(\hat{\mu}_i, \hat{\sigma}_i^2)$.	$1 \le i \le I$
17. $\hat{f}$	The equivalent flow of f with distribution denoted by $N(\hat{\mu}, \hat{\sigma}^2)$.	
	Equivalent traffic flow specification	
18. $\alpha_{i,j}$	QoS parameter of $\hat{f}_{i,j}$.	$1 \le i \le I$, $1 \le j \le J$
19. $\hat{\alpha}_i$	QoS parameter of $\hat{f}_i$.	$1 \le i \le I$
20. $\hat{c}_i$	The effective bandwidth of $\hat{f}_i$ (i.e., $\hat{c}_i = \hat{\mu}_i + \hat{\alpha}_i\hat{\sigma}_i$).	
21. $\overline{N_i}$	The estimated number of packets belonging to $\hat{f}_i$ arrived in one SI.	
22. $\overline{L_i}$	The weighted average packet size of $\hat{f}_i$.	

8. References

[1] IEEE 802.11 WG: IEEE Standard 802.11-2007, Part 11: Wireless LAN MAC and Physical Layer Specifications, Mar. 2007.

[2] IEEE Std. 802.11e-2005, Part 11: Wireless LAN medium access control and physical layer specifications Amendment 8: medium access control (MAC) quality of service enhancements, Nov. 2005.

[3] S. Mangold, S. Choi, G. R. Hiertz, et. All, “Analysis of IEEE 802.11e for QoS support in Wireless LANs,” *IEEE Wireless Commun.*, vol. 10, no. 6, pp. 40-50, Dec. 2003.

[4] P. Ansel, Q. Ni, and T. Turletti, “FHCF: A fair scheduling scheme for IEEE 802.11e WLAN,” *INRIA research report*, no. 4883, Jul. 2003.

[5] W. F. Fan, D. Y. Gao, D. H. K. Tsang and B. Bensaou, ”Admission Control for variable bit rate traffic in IEEE 802.11e WLANs,” in *Proc. IEEE LANMAN’04*, pp. 61-66, Apr. 2004.

[6] Y. W. Huang, T. H. Lee and J. R. Hsieh, ”Gaussian approximation based admission control for variable bit rate traffic in IEEE 802.11e WLANs,” in Proc. *IEEE WCNC’07*, pp. 3768-3773, Mar. 2007.

[7] T. H. Lee and Y. W. Huang, ”Effective transmission opportunity allocation

scheme for real-time variable bit rate traffic flows with different delay bounds," *IET Commun.*, vol. 2, no. 4, pp. 598-608, 2008.

[8] D. Gao, J. Cai and C. W. Chen, " Admission control based on rate-variance envelop for VBR traffic over IEEE 802.11e HCCA WLANs", *IEEE Trans. Vehicular Tech.*, vol. 57, no. 3, May, 2008.

[9] C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "An efficient cross layer scheduler for multimedia Traffic in Wireless Local Area Networks with IEEE 802.11e HCCA", *ACM Mobile Computing and Commun. Review*, vol. 11, no. 3, pp. 31-46, Jul. 2007.

[10] Y. Higuchi, A. Foronda, C. Ohta, M. Yoshimoto and Y. Okada, "Delay guarantee and service interval optimization for HCCA in IEEE 802.11e WLANs," in *Proc. IEEE WCNC'07*, pp. 2080-2085, Mar. 2007.

[11] M. M. Rashid, E. Hossain, V. K. Bharggava, "Controlled channel access scheduling for guaranteed QoS in IEEE 802.11e-based WLANs," *IEEE Trans. Wireless Commun.* vol. 7, no. 4, pp.1287-1297, Apr. 2008.

[12] A. A. Bourawy, N. A. AbuAli and H. S. Hassanein, "A selectivity function scheduler for IEEE 802.11e", in *Proc. IEEE ISCC'09*, pp. 950-955, Jul. 2009.

[13] J. J. Huang, Y. H. Chen and C.Y. Chang, "An MSI-based scheduler for IEEE 802.11e HCCA," in *Proc. IEEE VTC'09*, pp. 1-5, Sep. 2009.

[14] H. Luo and M. L. Shyu, “An optimized scheduling scheme to provide quality of service in 802.11e Wireless LANs”, in *Proc. IEEE ISM’09*, pp. 651-656, Dec. 2009.

[15] B. S. Kim, S. Kim, Y. Fang, and T.F. Wong, “Two-step multipolling MAC protocol for wireless LANs,” *IEEE J. Select. Areas in Commun.*, vol. 23, no. 6, pp. 1276-1286, June 2005.

[16] H. S. Kim and N. B. Shroff, “Loss probability calculations and asymptotic analysis for finite buffer multiplexers,” *IEEE/ACM Trans. Networking*, vol. 9, no. 6, pp. 755-768, Dec. 2001.

[17] L. Georgiadis, R. Guerin, A. Parekh, “Optimal multiplexing on a single link: delay and buffer requirements”, *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp.1518-1535, Sep. 1997.

[18] MPEG-4 and H.263 video traces for network performance evaluation, http://www.tkn.tu-berlin.de/research/trace/trace.html, Oct. 2006.