Admission Control in IEEE 802.11e Wireless LANs

Deyun Gao and Jianfei Cai, Nanyang Technological University, Singapore
King Ngi Ngan, Chinese University of Hong Kong, Hong Kong

Abstract

Although IEEE 802.11 based wireless local area networks have become more and more popular due to low cost and easy deployment, they can only provide best effort services and do not have quality of service supports for multimedia applications. Recently, a new standard, IEEE 802.11e, has been proposed, which introduces a so-called hybrid coordination function containing two medium access mechanisms: contention-based channel access and controlled channel access. In this article we first give a brief tutorial on the various MAC-layer QoS mechanisms provided by 802.11e. We show that the 802.11e standard provides a very powerful platform for QoS supports in WLANs. Then we provide an extensive survey of recent advances in admission control algorithms/protocols in IEEE 802.11e WLANs. Our survey covers the research work in admission control for both EDCA and HCCA. We show that the new MAC-layer QoS schemes and parameters provided in EDCA and HCCA can be well utilized to fulfill the requirements of admission control so that QoS for multimedia applications can be provided in WLANs. Last, we give a summary of the design of admission control in EDCA and HCCA, and point out the remaining challenges.

With the diminishing costs of electronic hardware, IEEE 802.11 based wireless local area networks (WLANs) have been massively deployed in public and residential places such as classrooms, airports, and apartments, and more and more devices and peripherals are integrated with WLAN access capability. On the other hand, with the increasing popularity of multimedia applications, quality of service (QoS) support in communication networks has become more and more important. QoS can be interpreted as the ability of a network to provide some consistent services for multimedia delivery. The Internet Engineering Task Force (IETF) has defined two different frameworks, integrated services (IntServ) [1] and differentiated services (DiffServ) [2], for Internet traffic with QoS. Compared to wired networks, to provide QoS in wireless networks is even more challenging since wireless networks have limited bandwidth, and radio channels are error-prone, affected by multipath, shadowing, interference, weather, and so on.

For WLANs, IEEE 802.11 is designed for best effort services. The 802.11 standard specifies two medium access control (MAC) mechanisms: the mandatory distributed coordination function (DCF) and the optional point coordination function (PCF) [3]. The lack of a built-in mechanism for supporting real-time services makes it very difficult to provide QoS guarantees for multimedia applications. To enhance QoS support in 802.11, the IEEE 802.11 working group is currently working on a new standard, known as the IEEE 802.11e [4–6], which introduces the so-called hybrid coordination function (HCF). HCF includes two medium access mechanisms: contention-based channel access and controlled channel access (includes polling). Contention-based channel access is referred to as enhanced distributed channel access (EDCA), and controlled channel access is referred as HCF controlled channel access (HCCA).

Although the 802.11e standard has defined QoS-enabled MAC mechanisms, how to apply these mechanisms to different QoS issues is not specified. Among various QoS issues, admission control is an important component for the provision of guaranteed QoS parameters. The purpose of admission control is to limit the amount of traffic admitted into a particular service class so that the QoS of the existing flows will not be degraded, while at the same time the medium resources can be maximally utilized. In this article we provide a survey of recent advances in admission control algorithms/protocols in IEEE 802.11e WLANs. Our survey covers the research work on admission control for both EDCA and HCCA. Our purpose is to study how the new QoS schemes and parameters provided in EDCA and HCCA can be well utilized to fulfill the requirements of admission control so that QoS for multimedia applications can be provided in WLANs.

The article is organized as follows. We give an overview of 802.11, introducing the legacy DCF and PCF schemes. We introduce the 802.11e QoS-enabled MAC mechanisms including EDCA, HCCA, and some other schemes. After that we survey recent research work on admission control for both EDCA and HCCA. Finally, we conclude this article with some discussions.

Priority	User priority in 802.1D	Access category (AC)	Designation (informative)
Lowest	1	AC[0]	Background
	2	AC[0]	Background
	0	AC[1]	Best effort
	3	AC[1]	Video
	4	AC[2]	Video
	5	AC[2]	Video
	6	AC[3]	Voice
Highest	7	AC[3]	Voice

■ Table 1. *The mapping between the user priorities in 802.1D and the access categories in 802.11E.*

An Overview of IEEE 802.11

802.11 refers to a family of specifications developed by the IEEE for WLAN technology, which operates at either the 2.4 GHz industrial, scientific, and medical (ISM) band or the 5 GHz unlicensed national information infrastructure (UNII) band. In 802.11 WLAN, the MAC layer defines the procedures for 802.11 stations to share a common radio channel. DCF is a contention-based access control scheme targeted at delivering classic data services, while PCF is a contention-free access control scheme targeted for time-bounded services. In practice, most 802.11 products in the market only support DCF. Note that in 802.11 there are two ways to organize stations of WLANs: the infrastructure and ad hoc modes. In this article we only consider the infrastructure mode.

Distributed Coordination Function

DCF works as a "listen-before-talk" scheme based on carrier sense multiple access with collision avoidance (CSMA/CA) where stations listen to the medium to determine when it is free. If a station that has packets to send senses the medium is busy, it will defer its transmission and initiate a backoff counter. The backoff counter is a uniformly distributed random number between 0 and contention window (CW). Once the station detects that the medium has been free for a duration of DCF interframe space (DIFS), it starts a backoff procedure (i.e., decrementing its backoff counter as long as the channel is idle). If the backoff counter has reduced to zero and the medium is still free, the station begins to transmit. If the medium becomes busy in the middle of the decrement, the station freezes its backoff counter, and resumes the countdown after deferring for a period of time, which is indicated by the so-called network allocation vector (NAV) stored in the winning station's packet header.

It is possible that two or more stations begin to transmit at the same time. In such a case, a collision occurs. Collisions are inferred by no acknowledgment (ACK) from the receiver. After a collision occurs, all the involved stations double their CWs (up to a maximum value, CW_{max}) and compete to gain control of the medium next time. If a station succeeds in channel access (inferred by the reception of an ACK), the station resets its CW to CW_{min}.

We can see that DCF does not provide QoS supports since all stations operate with the same channel access parameters and have the same medium access priority. There is no mechanism to differentiate different stations and different traffic.

Point Coordination Function

PCF provides contention-free transmission. In PCF time is divided into superframes. A superframe includes a contention period (CP), where DCF is used, and a contention-free period (CFP), where PCF is used. A superframe starts with a beacon management frame transmitted by the access point (AP), which acts as a point coordinator. The time used by the AP to generate beacon frames is called target beacon transmission time (TBTT), which is announced in the previous beacon frame. PCF uses the point interframe space (PIFS), which is longer than a short interframe space (SIFS) but shorter than DIFS, to provide point coordinators higher priority in medium access than DCF stations.

During the CFP, the AP polls its associated stations according to a predetermined order called *polling list* (usually in a round-robin manner). No station is allowed to transmit unless it is polled. If there is no pending transmission in a polled station, the response is a null frame containing no payload. The CFP ends when the AP sends a CF-end message. If the CFP terminates before all stations have been polled, the polling list will be resumed at the next CFP cycle from the previous stopping point. If the AP receives no response from a polled station after waiting for a PIFS, it will poll the next station or end the CFP. In this way, no idle period longer than a PIFS occurs during a CFP.

Compared to DCF, PCF is more complicated and requires central control. As for supporting QoS, PCF still has some problems [7, 8]. For example, PCF has the unpredictable beacon delay problem: a beacon has to be delayed if there is an unfinished DCF frame at the end of the previous superframe. Another problem is that it is very difficult to predict the transmission time of a polled station because the polled station can transmit a frame of any length between 0 and the size of the maximum MAC service data unit (MSDU).

IEEE 802.11E QoS Mechanisms

Due to the limitations of DCF and PCF, the 802.11e defines a single coordination function, HCF, which combines the functions of both DCF and PCF for QoS data transmission. In 802.11e a superframe still consists of the two phases of operations, CP and CFP. EDCA is only used in the CP, while HCCA can be used in both phases. The major benefits offered by the 802.11e standard are:

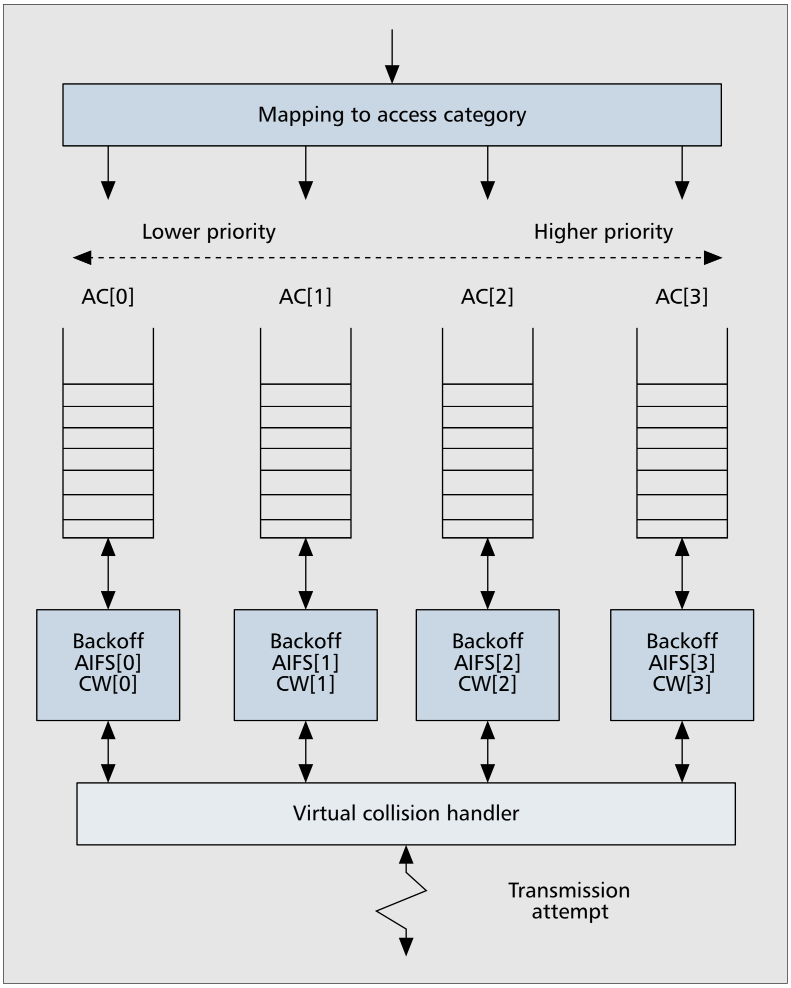
- Reducing the latency through prioritizing different types of traffic packets
- Enabling APs to allocate resources based on data rate and latency requirements from each individual station
- Improving wireless bandwidth efficiency and reducing packet overheads

HCF Contention-Based Channel Access

In EDCA, the QoS support is realized through introducing multiple access categories (ACs) in each QoS station (QSTA). EDCA defines four ACs, and different ACs have different priorities, servicing different types of traffic. Table 1 [4] shows the mapping between the user priorities (UPs) specified in IEEE 802.1D [9] and the ACs in 802.11e.

As shown in Fig. 1, each AC is an enhanced variant of DCF that contends for transmission opportunity (TXOP) using AC-specified channel access parameters from the EDCA parameter set, which includes:

- Minimal CW value for a given AC ($CW_{min}[AC]$): CW_{min} can be different for different ACs. Assigning smaller values of CW_{min} to high priority classes can ensure that high-priority classes obtain more TXOPs than low-priority ones.
- Maximal CW value for a given AC ($CW_{max}[AC]$): Similar to CW_{min}, CW_{max} is also on a per AC basis.
- Arbitration interframe space ($AIFS[AC]$): Each AC starts its backoff procedure after the channel is idle for a period of $AIFS[AC]$ instead of DIFS. The $AIFS[AC]$ for a given AC should be equal to an SIFS plus multiple time slots (i.e., $AIFS[AC] = aSIFSTime + AIFSN[AC]*aSlotTime$). Considering $DIFS = aSIFSTime + 2 * aSlotTime$ in legacy 802.11, $AIFSN[AC]$ is typically set to not less than 2 such that the shortest waiting time is DIFS.
- *TXOPlimit*[*AC*]: TXOPs obtained via EDCA are referred as EDCA-TXOPs. During an EDCA-TXOP, a station may be allowed to transmit multiple data frames from the same AC with a SIFS gap between an ACK and the subsequent data frame transmission. *TXOPlimit*[*AC*] gives the limit for such a consecutive transmission.
- Virtual collision: If the backoff counters of two or more collocated ACs in one station elapse at the same time, a scheduler inside the station treats the event as a virtual collision. The TXOP is given to the AC with the highest priority among the "colliding" ACs, and the other colliding ACs defer and try again later as if the collision occurred in the real medium.



■ Figure 1. *The virtual backoff of the four access categories.*

HCF Controlled Channel Access

Although EDCA improves the legacy DCF, it is not sufficient to provide effective traffic protection and QoS guarantees, especially under high traffic loads. Here comes the need for the polling-based medium access mechanism, HCCA.

Similar to the legacy PCF, HCCA provides polled access to the wireless medium. In particular, HCCA uses a QoS-aware hybrid coordinator (HC), which is typically located at the QoS access point (QAP) in infrastructure WLANs. HC uses PIFS to gain control of the channel and then allocates TXOPs to QSTAs, which are referred as HCCA TXOPs or polled TXOPs. Unlike PCF, HCCA can poll the QSTAs during contention periods (CPs), and HCCA takes into account QSTAs' specific flow requirements in packet scheduling. Figure 2 illustrates the different periods under HCCA. Note that the controlled access phase (CAP) is defined as the time period when HC maintains control of the medium. It can be seen that CAPs consist of not only CFPs but also parts of CPs.

After grabbing the channel, the HC polls QSTAs in turn according to its polling list. In order to be included in the polling list of the HC, a QSTA must send a QoS reservation request using the special QoS management frame, and each individual flow needs one particular reservation request. Figure 3 depicts a common frame format for carrying traffic specification (TSPEC) parameters. The major TSPEC parameters [4, 10] include:

- **Mean data rate** (ρ): the average bit rate for packet transmission, in bits per second
- **Delay bound (D)**: the maximum delay allowed to transport a packet across the wireless interface (including queuing delay), in milliseconds
- **Maximum service interval** (*SImax*): the maximum time allowed between neighbor TXOPs allocated to the same station, in microseconds
- **Nominal MSDU size (L)**: the nominal size of a packet, in octets
- **Minimum PHY rate (R)**: the minimum physical bit rate assumed by the scheduler for calculating transmission time, in bits per second

For the definitions of other parameters, details can be found in the 802.11e draft [4].

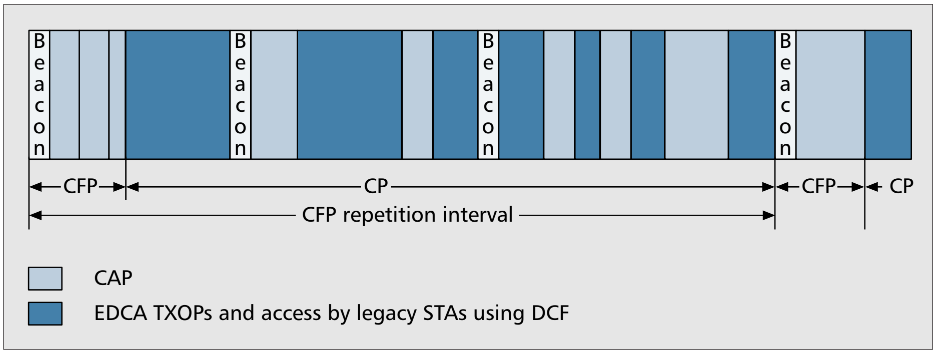
Note that, although the QAP decision is based on the characteristics of individual flows, the HCCA TXOPs are actually assigned on a per QSTA basis instead of per flow, and each QSTA is then responsible for allocating the TXOPs to its individual flows.

Other QoS Mechanisms

In addition to EDCA and HCCA, the 802.11e draft also provides some other MAC mechanisms for QoS enhancements.

Block acknowledgment (block ACK): Block ACK can be initiated through a setup and negotiation process between a QSTA and the QAP. Once the block ACK has been established, multiple QoS data frames are transmitted in a contention-free burst with an SIFS interval between adjacent frames. This mechanism helps to reduce the bandwidth overheads imposed by the conventional ACK mechanism, which requires an individual ACK to every successful data frame.

Direct link protocol (DLP): DLP allows two stations associated with the same QAP to directly



■ Figure 2. *The CAP/CFP/CP periods.*

transmit data to each other. This facilitates efficient use of the transmission medium as transmissions do not need to be directed through the QAP any more.

No acknowledgment (no ACK): For certain applications, 802.11e allows no ACK to be sent. This feature is very useful for applications that have stringent delay requirements but can tolerate a significant amount of packet loss (e.g., voice over IP).

Piggyback: 802.11e allows data to be sent "piggybacked" on polls and ACKs to reduce overhead. This feature can improve overall network performance.

Through combining the various MAC QoS mechanisms in 802.11 together, it is possible to provide QoS and efficiently utilize medium resources at the same time.

Admission Control for EDCA

Similar to DCF, EDCA is very likely to be the dominant channel access mechanism in WLANs because it is a distributed MAC scheme and easy to implement. In the past few years, we have seen much research work focusing on admission control in EDCA. Basically, the existing EDCA admission control schemes can be classified into two categories: measurement-based and model-based. In measurement-based schemes, admission control decisions are made based on continuously measured network conditions such as throughput and delay. On the other hand, the model-based schemes construct certain performance metrics to evaluate the status of the network.

Measurement-Based Admission Control

Distributed Admission Control (DAC) [11, 12] — DAC was proposed by the 802.11e working group to protect active QoS flows. Even though it is not supported in the latest draft, it is a good starting point to study admission control mechanisms in EDCA.

In the DAC, via beacons the QAP announces the transmission budget, which is the additional amount of time available for each AC during the next beacon interval. In order to calculate transmission budget, the QAP needs to measure the amount of time occupied by the transmission of each AC during each beacon interval. Then the QAP computes the transmission budget for each AC by subtracting the occupied time from the transmission limit of this AC. Each station determines an internal transmission limit per AC for each beacon interval based on the successfully used transmission time during the previous beacon period and the transmission budget announced from the QAP. When the transmission budget for an AC is depleted, a new flow will not be able to obtain any transmission time, and existing flows will not be able to increase their transmission time either.

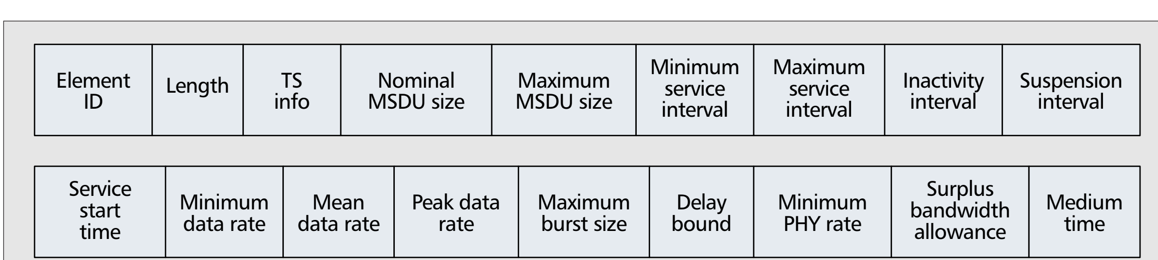
One problem of the DAC scheme is that it is difficult to avoid network performance vibration because a station always adjusts its transmission parameters at every beacon interval. Another shortcoming of the DAC scheme is that it can only protect existing flows when the traffic load is not very heavy. In addition, this scheme does not provide direct relationships between those TXOP parameters and the QoS requirements from applications.

Two-Level Protection and Guarantee Mechanism [13] — Based on the DAC scheme, the authors further proposed a two-level protection and guarantee scheme. The purpose of the first-level protection is to protect each existing voice or video flow from new and other existing QoS flows, while the purpose of the second-level protection is to protect the existing QoS flows from best effort traffic.

In particular, two enhancements, *tried-and-known* and *early-protection*, are introduced in the original DAC scheme. In the tried-and-known mechanism, a new voice/video flow is first accepted tentatively, and then tries to measure throughput and delay performance for some beacon intervals. If the average throughput and/or delay do not meet reasonable requirements, the flow will kill or reject itself. In the early-protection mechanism, when the budget is below a certain threshold, new flows are not allowed to enter. This first-level protection performs well in protecting each individual existing QoS flow from new and other existing QoS flows.

However, too many best effort data transmissions can also degrade the existing QoS flows since many collisions might occur. Therefore, the second-level protection (i.e., dynamically controlling the EDCA parameters) is introduced. The basic idea is to increase the initial contention window size and interframe space for best-effort traffic when the number of active stations is large. In this way, the number of collisions can be kept relatively small.

Because this two-level protection scheme is based on the DAC, it also has the problems of performance oscillation and lack of direct QoS relationships with applications. Moreover,



■ Figure 3. *A common management frame format for traffic specification.*

it introduces more adjustable parameters, and finding the optimized parameters is not a trivial task.

Virtual MAC and Virtual Source Algorithms [14, 15] — The basic idea of the virtual MAC and virtual source algorithms is to virtually run the applications and MAC processes in order to measure the achievable service qualities. Based on the virtually measured service qualities, QSTAs determine whether a new flow should be admitted or not.

The virtual MAC (VMAC) algorithm operates in parallel to the real MAC in a station, and it handles "virtual packets" instead of real packets. The VMAC schedules virtual packets on the radio channel in the same way as real packets. However, unlike the case of real packets, when the VMAC decides to send a virtual packet, it does not transmit anything but estimates the probability of collision if the virtual packet were "really" sent. When a collision is "detected," the VMAC enters a backoff procedure, just as a real MAC would do.

The virtual source (VS) algorithm consists of a virtual application, an interface queue, and the VMAC. The virtual application generates virtual packets like a real application (e.g., generating virtual voice packets at a constant rate). Packets are timestamped and placed in a virtual buffer. After a virtual packet has been processed in the VMAC, the total delay is calculated by comparing the current time with the timestamp stored in the packet.

The advantage of these virtual algorithms is that they do not cost any channel bandwidth. However, they need a lot of extra processing in each mobile host. Also, the main criteria for the admission decision are based on delay and collision estimations. No achievable throughput information is available.

Threshold-Based Admission Control [16] — In this scheme, each station needs to measure the traffic condition on the wireless link. Depending on how the traffic condition is measured and computed, the admission control can be implemented in the following two ways:

- **Using relative occupied bandwidth**: T_{Busy} is defined as the amount of time when the wireless medium is busy. The relative occupied bandwidth is defined as $B_{occu} = T_{Busy}/T \times 100$ percent, where T is a fixed sampling period. Let B_{lo}, B_{up} be the given lower and upper thresholds for B_{occu}. Let an active AC denote an admitted AC and an inactive AC denote an AC refused admission. We summarize the criteria to admit data flows as:
  - $B_{occu} \le B_{lo}$: Admit the inactive AC with the highest priority during the next period of T.
  - $B_{lo} \le B_{occu} \le B_{up}$: No action taken.
  - $B_{occu} \ge B_{up}$: Stop the transmission of the lowest active AC during the next period of T.
- **Using average collision**: In this case, instead of using the relative occupied bandwidth, a new variable, average collision ratio, is employed for admission control. The average collision ratio is defined as $R_c = N_c/N_t$, where N_c is the number of collisions that have occurred and N_t is the total number of transmissions. Similarly, there are two thresholds: the lower threshold R_{lo} and the upper threshold R_{up}. Except using the new parameters R_c, R_{lo}, and R_{up}, the criteria for admission control are the same as in the case of using relative occupied bandwidth.

Although this threshold-based admission control is very easy to implement, the threshold values are difficult to set. In addition, since the transmission of data flows will be stopped when network resources are unavailable or resumed when network resources are available, there is no way to guarantee the instantaneous QoS metrics.

Harmonica [17] — In the HARMONICA scheme, the AP periodically samples the link layer quality indicator (LQI) parameters, which include drop rate, link layer end-to-end delay, and throughput, for each traffic class. Two adaptation algorithms over different timescales are employed to select the channel access parameters that best match the QoS requirements of each traffic class and the current channel contention level. In particular, one adaptation algorithm adjusts the relative differences between the channel access parameters of different classes for the purpose of QoS guarantee. The other adaptation algorithm synchronously adapts the channel access parameters of all the classes (increase all or decrease all) to achieve high channel utilization.

HARMONICA employs a simple and flexible admission control mechanism to avoid congestion. Whenever a new real-time application requires admission, the HARMONICA will select a traffic class i that best matches its QoS requirement and then execute an admission control process. The decision of admission control is based on the throughput requirement ($Req_{throughput}$) of the flow and the monitored LQI parameters. The key idea is to check whether it is possible to squeeze some bandwidth out of the current throughput for the best effort class ($BE_{throughput}$) on the condition of guaranteeing a minimal bandwidth (BE_{Min}) for the best effort traffic class. In particular, in order to admit a new QoS flow, three requirements need to be satisfied:

- The relative adaptation has reached a stable state.
- $BE_{throughput} - Req_{throughput} \ge BE_{Min}$.
- The bandwidth in $BE_{throughput}$ can be "translated" into class i without loss.

The HARMONICA scheme shows that through dynamically adjusting channel access parameters, it is possible to simultaneously match the QoS requirements, maximally utilize network resources, and guarantee a minimal bandwidth for best effort traffic. However, how to find the optimal increment or decrement of the channel access parameters is still a challenging problem. If the chosen increment or decrement is too large, the systems will oscillate. On the other hand, if the chosen increment or decrement is too small, the system will take a long time to reach the optimal status.

Model-Based Admission Control

Markov Chain Model-Based Admission Control [18] — In this scheme, admission control is performed based on the predicted achievable throughput for each flow, which is calculated by

$$S_i = \frac{P_{si} \times E[P]}{P_c \times T_c + P_{idle} \times aSlotTime + P_s \times T_s}. \qquad (1)$$

P_{si} is the probability of a successful transmission for flow i, which is given by the product of the probability that flow i is transmitting and another probability that no other flow is transmitting. On the other hand, if multiple flows transmit in the same time slot, a collision occurs. P_c, P_s, and P_{idle} are the overall collision probability, overall successful transmission probability, and overall idle (no transmission) probability, respectively. T_c and T_s are the collision time and successful transmission time, respectively. $E[P]$ is the data payload.

From the definitions of P_{si}, P_c, P_s, and P_{idle}, we can see that in order to calculate these probabilities, we need to first find a way to obtain the transmission probability for a flow. Luckily, based on the two-state Markov Chain model proposed in [19], the transmission probability for flow i can be derived as

$$P_{ti} = \frac{2(1-2p_i)}{(1-2p_i)(W+1)+p_i W(1-(2p_i)^b)}, \quad (2)$$

where p_i is the long-term collision probability for flow i, W is the CW_{min} size for flow i, and b is the maximum backoff stage. p_i can be obtained by using a counter for each active flow to keep track of the collision rate.

There are several problems in this algorithm. The first one is that the analytical model is derived under saturation conditions, where each station always has packets to transmit. In fact, this is not always true since in practice we often experience nonsaturation conditions. Another problem is that the analytical model in [18, 19] does not take account of virtual collision between different AC queues in one station. Thus, if there is more than one flow in one station, the analytical model is not accurate at all.

Contention-Window-Based Admission Control [20] — The key idea of this scheme is to adjust the CW values for different stations so that the goals of admission control can be fulfilled. In particular, suppose an IEEE 802.11e WLAN is operating with a CW set $\{CW_1, \ldots, CW_n\}$ that meets the throughput requirements $\{R_i, \ldots, R_n\}$ for all stations. Let r_i denote the actual throughput experienced by station i, where $r_i \geq R_i$. When a new station $(n + 1)$ wants to join the network with a throughput requirement of R_{n+1}, based on the proposed analytical model in [20], a new CW set $\{CW'_1, \ldots, CW'_n, CW'_{n+1}\}$ will first be calculated and then used to compute the throughput. If the resulting throughput meets the requirements, station $(n + 1)$ is accepted and the new CW set is distributed to all the stations. Otherwise, station $(n + 1)$ is rejected.

Similar to the Markov-chain-based scheme [18], this CW-based scheme also has the problems of not considering nonsaturation conditions and virtual collision. In addition, only using CW to adjust throughput is somewhat limited in terms of fully utilizing the capability of 802.11e.

Admission Control for HCCA

Unlike EDCA, there is not much research work on the admission control issue in HCCA. This is mainly due to the centralized control of HCCA, which results in the deterministic nature of admission control in HCCA. Thus, in terms of research, the admission control of HCCA is not as challenging as that of EDCA. Another reason could be that the distributed MAC mechanism is much more popular than the centralized mechanism in practice. In the following, we introduce a reference admission control scheme in 802.11e and our previous work in this area.

Reference Scheme

A reference admission control algorithm is developed in the 802.11e draft [4, 10]. In particular, admission control is based on a simple scheduler, which uses the mandatory set of TSPEC parameters to generate a schedule. The mandatory set of TSPEC parameters are mean data rate (ρ), nominal MSDU size (L), and maximum service interval (SI_{max}) or delay bound (D). When a new flow requests for admission, the admission control process is preformed in three steps:

- Step 1: The admission control unit (ACU) calculates the number of MSDUs that arrive at the mean data rate during the scheduled service interval SI as

$$N_i = \left\lceil \frac{SI \times \rho_i}{L_i} \right\rceil. \quad (3)$$

Note that the scheduled service interval SI must be a number lower than the minimum value of all the maximum service intervals for all the admitted flows, and must also be a submultiple of the beacon interval.

- Step 2: For a flow i, TXOPi is calculated as

$$TXOP_i = \max\left(\frac{N_i \times L_i}{R_i} + O, \frac{M}{R_i} + O\right), \quad (4)$$

where R_i is the minimum physical transmission rate, M is the maximum size of an MSDU, and O is the overhead in time units. The overhead includes interframe spaces, ACKs, and so on.

- Step 3: Assuming there are k admitted flows, a new flow $k + 1$ is accepted if it satisfies

$$\frac{TXOP_{k+1}}{SI} + \sum_{i=1}^{k} \frac{TXOP_i}{SI} \leq \frac{T - T_{CP}}{T}, \quad (5)$$

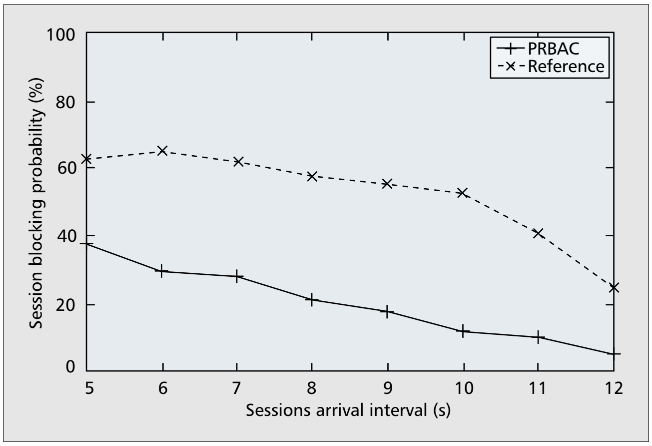
where T is the beacon interval and T_{CP} is the time for EDCA traffic.

Physical-Rate-Based Admission Control

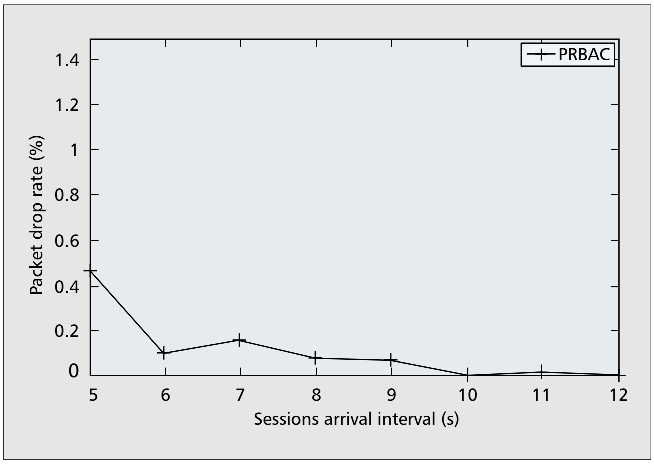
The above reference design is somewhat inefficient because it is implemented based on the minimum physical rate. The actual physical rate of a station is quite different from the minimum physical rate most of the time. Therefore, in our previous work [21] we proposed a more efficient admission control algorithm named physical rate-based admission control (PRBAC), which considers physical rate variance due to station mobility and wireless channel characteristics. Most WLAN products adjust their physical rates according to the estimated wireless channel conditions [22, 23]. Stations use lower rates when they are far away from the AP and use higher rates when they move nearer to the AP. Low rates could also be selected in the cases of high path loss, high background noise, and extreme multipath effects in order to increase the robustness. In our proposed PRBAC, the basic idea is to use the long-term average physical rates for admission control, and at the same time use the instantaneous physical rates to distribute TXOPs to individual stations. In this way, our algorithm can admit more flows than the reference scheme, while the performance of each individual station is not degraded too much because the instantaneous physical rate is used to calculate TXOPs.

Note that in our proposed scheme, sometimes we have to drop packets. For example, when a lot of stations move away from the AP, it is very likely that the network resources will become insufficient; thus, dropping packets is unavoidable. We handle such a problem by randomly selecting a QoS session and reducing its TXOPs by the amount of time for one packet. The same process is repeated continuously until Eq. 5 is satisfied.

We have compared our proposed scheme with the reference scheme by simulation. In our simulation, 51 stations form a QBSS with one station being the QAP. We assume that the arrival of QoS session requests from QSTAs to the QAP is a Poisson process, and one QSTA has at most one QoS session at a time. In addition, each QSTA transfers background data traffic to the QAP, which employs the EDCA mechanism to access the medium. All the QSTAs move in a pattern according to the "random waypoint" mobility model [24]. The QoS sessions arrival interval is changed from 5 s to 12 s in the simulation. Figures 4 and 5 show the session blocking probability and packet drop probability for our proposed scheme and the reference scheme. We can see that our pro-



■ Figure 4. *The QoS session blocking probability for PRBAC and the reference scheme.*



■ Figure 5. *The packet drop rate for PRBAC.*

posed scheme utilizes network resources much better than the reference scheme with only small performance degradation. More details can be found in [21].

Admission Control for VBR Traffic

The above two schemes only consider the mean sending rate and the mean packet size of a flow. However, for variable bit rate (VBR) traffic, the instantaneous sending rate and the packet size are usually quite different from the corresponding mean values. Thus, a new admission control scheme for VBR traffic was further proposed in [25]. The key idea is to introduce a new variable, effective TXOP (T_e), to replace TXOP in Eq. 5. T_e is defined as the necessary TXOPs which can statistically guarantee that the packet loss probability is less than a threshold. In particular, there are two cases in calculating the packet loss probability.

VBR traffic with constant packet size: For constant packet size, only the packet arrival rate is varying, and the packet loss rate can be expressed as the mean number of packets lost during *SI* over the mean number of packets arriving during *SI*. Thus, the packet loss rate only depends on the probability distribution of the number of arrived packets.

VBR traffic with variable packet size: For variable packet size, both the packet arrival rate and packet size vary. The packet loss rate should be expressed in terms of packet transmission time rather than number of packets, i.e., the mean transmission time of the lost packets during *SI* over the mean transmission time of all the arrived packets during *SI*.

Given a desired packet loss probability, the effective TXOP duration of a newly arrived VBR flow can be inversely derived from the packet loss rate expression. Then the same procedure as in the reference scheme is applied for admission control, except we use the effective TXOP durations in Eq. 5.

Besides our proposed admission control scheme for VBR traffic, the work in [26] also takes the characteristics of VBR traffic into consideration. In [26] a fair scheduling algorithm named fair HCF (FHCF) was proposed to replace the simple round-robin scheduler in HCF. In particular, the FHCF scheme consists of two types of schedulers: QAP and node. The QAP scheduler estimates the queue length for each QSTA before the next *SI*. Based on these estimated queue lengths, the QAP adapts the computation of TXOPs accordingly. The node scheduler is to redistribute the unused time among its different traffic streams. We believe this FHCF scheme can be combined with the scheme in [25] to enable dynamic admission control and traffic scheduling for VBR traffic.

Conclusion

In this article we have briefly introduced many MAC-layer QoS mechanisms provided in the upcoming 802.11e standard. These MAC-layer QoS mechanisms make the 802.11e standard a very powerful platform to support QoS in WLANs for multimedia applications. After that, we have surveyed various admission control schemes for both EDCA and HCCA. In particular, for EDCA we have classified admission control into two categories: measurement-based and model-based. Measurement-based schemes are usually effective and simple to implement in practice. However, without a theoretical foundation, it is very difficult to achieve overall optimization. On the contrary, model-based admission control schemes are based on some analytical models, and it is possible to optimize the entire system. However, those analytical models are usually derived based on some unrealistic assumptions such as error-free physical channels for the purpose of simplifying the derivation. Thus, the optimal solutions obtained by the analytical models might not be suitable for practical systems. We believe a joint measurement- and model-based approach will be a good method of admission control in EDCA.

As pointed out earlier, for HCCA there is not much research work on the admission control issue. This is mainly due to the centralized control of HCCA, which results in the deterministic nature of admission control in HCCA. Thus, in terms of research, admission control in HCCA is not as challenging as that in EDCA. For HCCA we have described the reference scheme of admission control in the 802.11e draft, our previously proposed physical-rate-based admission control, and the admission control schemes for VBR traffic.

Similar to the relationship between DCF and PCF, EDCA performs better under light traffic load due to less overhead, while HCCA performs better under heavy traffic load due to less collision. Thus, EDCA and HCCA cannot replace each other; they can only complement to each other. As for the design of admission control, there are some fundamental differences between EDCA and HCCA, such as:

- HCCA provides deterministic QoS performance for applications with admission control, while EDCA only provides statistical QoS performance. This is because HCCA is contention-free and EDCA is contention-based.
- Admission control for HCCA can only be used in infrastructure mode, while many admission control schemes in EDCA can be used in both infrastructure and ad hoc modes.

Although many admission control schemes have been proposed for EDCA and HCCA, a complete and comprehensive solution is still not available. There are many challenges remaining. For instance, in HCCA it is not clear how to optimally select effective TXOPs to trade off between flow QoS and network utilization. Another challenge is how to trade off between HCCA and EDCA in a mixed HCCA and EDCA scenario. For EDCA, the major challenge is how to optimally map the QoS requirements from applications into the channel access parameters. With the trend toward all-IP networks in the future, it is even more challenging to develop admission control schemes under heterogeneous wireless networks.

References

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

[2] S. Blake *et al.*, "An Architecture for Differentiated Service," RFC 2475, Dec. 1998.

[3] IEEE Std. 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.

[4] IEEE P802.11e/D6.0, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements," Nov. 2003.

[5] X. Yang, "IEEE 802.11e: QoS Provisioning at the MAC Layer," *IEEE Wireless Commun. Mag.*, vol. 11, no. 3, Mar. 2004, pp. 72–79.

[6] S. Mangold *et al.*, "Analysis of IEEE 802.11e for QoS Support in Wireless LANs," *IEEE Wireless Commun. Mag.*, vol. 10, no. 6, 2003, pp. 40–50.

[7] A. Lindgern, A. Almquist, and O. Schelen, "Evaluation of Quality of Service of Service Schemes for IEEE 802.11 wireless LANs," *Proc. 26th Annual IEEE Conf. Local Comp. Networks*, FL, Nov. 2001, pp. 348–51.

[8] S. Mangold *et al.*, "IEEE 802.11 Wireless LAN for Quality of Service," *Proc. Euro. Wireless*, Florence, Italy, Feb. 2002.

[9] IEEE Std. 802.1D, "Media Access Control (MAC) Bridges," 2004.

[10] A. Grilo, M. Macedo, and M. Nunes, "A Scheduling Algorithm for QoS Support in IEEE 802.11e Networks," *IEEE Wireless Commun.*, vol. 10, no. 3, June 2003, pp. 36–43.

[11] Y. Xiao and H. Li, "Evaluation of Distributed Admission Control for the IEEE 802.11e EDCA," *IEEE Commun. Mag.*, vol. 42, no. 9, 2004, pp. S20–S24.

[12] Y. Xiao and H. Li, "Voice and Video Transmissions with Global Data Parameter Control for the IEEE 802.11e Enhance Distributed Channel Access," *IEEE Trans. Parallel Distrib. Sys.*, vol. 15, no. 11, 2004, pp. 1041–53.

[13] Y. Xiao, H. Li, and S. Choi, "Protection and Guarantee for Voice and Video Traffic in IEEE 802.11e Wireless LANs," *Proc. IEEE INFOCOM '04*, vol. 3, Hong Kong, Mar. 2004, pp. 2152–62.

[14] M. Barry, A. T. Campbell, and A. Veres, "Distributed Control Algorithms for Service Differentiation in Wireless Packet Networks," *Proc. IEEE INFOCOM '01*, vol. 1, Anchorage, AK, 2001, pp. 582–90.

[15] A. Veres *et al.*, "Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control," *IEEE JSAC*, vol. 19, no. 10, 2001, pp. 2081–93.

[16] D. Gu and J. Zhang, "A New Measurement-based Admission Control Method for IEEE 802.11 Wireless Local Area Networks," Mitsubishi Elec. Research Lab., Tech. rep. TR-2003-122, Oct. 2003.

[17] L. Zhang and S. Zeadally, "HARMONICA: Enhanced QoS Support with Admission Control for IEEE 802.11 Contention-based Access," *Proc. IEEE RTAS '04*, Toronto, Canada, May 2004, pp. 64–71.

[18] D. Pong and T. Moors, "Call Admission Control for IEEE 802.11 Contention Access Mechanism," *Proc. IEEE GLOBECOM'03*, vol. 1, San Francisco, CA, Dec. 2003, pp. 174–78.

[19] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE JSAC*, vol. 18, no. 3, 2000, pp. 535–47.

[20] A. Banchs, X. Perez-Costa, and D. Qiao, "Providing Throughput Guarantees in IEEE 802.11e Wireless LANs," *Proc. 18th Int'l. Teletraffic Cong.*, Berlin, Germany, Sept. 2003.

[21] D. Gao, J. Cai, and L. Zhang, "Physical Rate-Based Admission Control for HCCA in IEEE 802.11e WLANs," *Proc. IEEE 19th Int'l. Conf. Adv. Info. Net. and Apps.*, Taiwan, Mar. 2005.

[22] A. Kamerman and L. Monteban, "WaveLAN-II: A High-Performance Wireless LAN for the Unlicensed Band," *Bell Lab Tech. J.*, Summer 1997, pp. 118–33.

[23] M. Lacage, M. H. Manshaei, and T. Turletti, "IEEE 802.11 Rate Adaptation: A Practical Approach," *Proc. 7th ACM Int'l. Symp. Modeling, Analysis and Simulation of Wireless and Mobile Sys.*, Venice, Italy, 2004, pp. 126–34.

[24] J. Broch *et al.*, "A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols," *Proc. ACM/IEEE MOBICOM '98*, Dallas, TX, Oct. 1998.

[25] W. F. Fan *et al.*, "Admission Control for Variable Bit Rate Traffic in IEEE 802.11e WLANs," *Proc. IEEE LANMAN '04*, Mill Valley, CA, 2004, pp. 61–6.

[26] Q. Ni, P. Ansel, and T. Turletti, "A Fast Scheduling Scheme for HOF," IEEE 802.11e WG doc. 802.11-03-0577-01-000e, July 2003.

Biographies

DEYUN GAO (dygao@ntu.edu.sg) received B.Eng. and M.Eng. degrees in electrical engineering and a Ph.D. degree in computer science from Tianjin University, China, in 1994, 1999, and 2002, respectively. From 2002 to 2003 he worked as a research associate with the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Kowloon. He is currently a research fellow in the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests are in the area of mobile and wireless Internet with emphasis on QoS guarantee, multimedia traffic delivery, and MAC protocol, and next-generation wireless networks.

JIANFEI CAI received his B.S. degree in electrical engineering from Nanjing University of Aeronautics and Astronautics in 1994, his M.S. degree in electrical engineering from Southeast University in 1997, and his Ph.D degree in electrical engineering from the University of Missouri-Columbia in 2002. He is currently an assistant professor with the School of Computer Engineering, Nanyang Technological University. His research interests include image/video compression, multimedia networking, and wireless communications.

KING N. NGAN [M'79, SM'91, F'00] received a Ph.D. degree in electrical engineering from Loughborough University of Technology, United Kingdom. He is a chair professor with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin. Previously, he was a full professor at Nanyang Technological University, Singapore, and the University of Western Australia, Crawley. He is an Associate Editor for *Journal on Visual Communications and Image Representation* and an Area Editor of *EURASIP Journal of Image Communication* and *Journal of Applied Signal Processing*. He has chaired a number of prestigious international conferences on video signal processing and communications, and has served on the advisory and technical committees of numerous professional organizations. He has published extensively, including three authored books, five edited volumes, and over 200 refereed technical papers in the areas of image/video coding and communications. He is a Fellow of the IEE (United Kingdom) and a Fellow of IEAust (Australia). He was an Associate Editor of *IEEE Transactions on Circuits and Systems for Video Technology*.