QoS Provisioning for Wireless LANs With Multi-Beam Access Point

Zi-Tsan Chou, *Member, IEEE,* Cong-Qi Huang, and J. Morris Chang, *Senior Member, IEEE*

Abstract—Recently, the integration of smart antenna technology into existing wireless local area networks (WLANs) has been one of the hot spots of research work. In this paper, we design an IEEE 802.11-compliant medium access control (MAC) protocol, named M-HCCA, that fully takes advantage of multi-beam smart antennas equipped at the access point (AP) to not only boost the overall capacity of a WLAN, but also support quality-of-service (QoS) and power conservation for individual mobile users. Specifically, M-HCCA has the following attractive features: (i) since being a polling-based MAC scheme, M-HCCA can innately conquer the problems induced by carrier sensing or directional signals, including beam-synchronization constraint, receiver blocking problem, and unnecessary defer problem; (ii) M-HCCA achieves high real-time throughput by adaptively adjusting the sector configuration to quickly resolve contention/collision and to increase data transmission parallelism; (iii) M-HCCA employs beam-location-aware polling scheduling to not only solve the beam-overlapping problem and back/side-lobe problem, but also let real-time stations save as much energy as possible; (iv) M-HCCA adopts the mobile-assisted admission control technique such that the AP can admit as many newly streams as possible while not violating QoS guarantees made to already-admitted streams; (v) M-HCCA offers a location updating mechanism to promptly renew the beam-location information of a non-responsive station such that the miss-hit problem can be effectively alleviated. Extensive simulation results show that, in terms of throughput, real-time throughput, and energy throughput, M-HCCA significantly outperforms existing protocols even in uneven station distribution, imperfect beam-forming, and high mobility environments.

Index Terms—Medium access control (MAC), multimedia, power management, quality of service (QoS), switched multi-beam antenna, and wireless local area network (WLAN)

1 Introduction

A wireless local area network (WLAN) typically consists of an access point (AP) and a finite set of mobile stations. Since the AP is generally more powerful and less physical constraint than mobile stations, it is of great interest to consider the use of *smart antennas* equipped at the AP to boost the network throughput by exploiting spatial reuse. According to [16], the existing smart antennas could be broadly classified into three categories: switched multi-beam antennas (SMBAs), adaptive array antennas, and multiple-input-multiple-output (MIMO) links. Clearly, each of these antenna technologies has its pros and cons. In this paper, we focus on SMBAs since they are relatively simple, commercial available, and have been deployed (for example, in Taipei, Taiwan) [9], [11], [14], [19]. The superior capabilities of smart antennas, however, can be leveraged only through appropriately designed higher layer network protocols, including at the *medium access control* (MAC) layer.

1.1 MAC Design Challenges for Multi-beam Antennas

Our considered multi-beam smart antenna model follows the assumptions of [17], [19]. Specifically, as shown in Fig. 1, the antenna system at the AP consists of M sectors. Each sector S_i contains $n_i \geq 1$ narrow beams, where $\sum_{i=0}^{M-1} n_i = N$ and $N/M = \omega$ is a positive integer. Each beam b_j has a beamwidth of about $360^\circ/N$ degrees, where $0 \leq j \leq N-1$. Note that the authors of [19] indicated that if one sector consists of only one wide-beam antenna, the front to back lobe ratio will be low, thus causing significant interference to other sectors. Furthermore, we assume that each sector is equipped with one individual transceiver. Hence we can treat multiple narrow-beam antennas in each sector as one logical antenna; besides, in each sector, at most one mobile station can communicate with the AP at the same time. To take the backward compatibility into account, we assume that mobile stations use omnidirectional antennas and all beams at the AP operate in the same frequency band. Referring to Fig. 1, since the AP has three sectors and stations A, B, and C are located in different sectors, the AP can concurrently send different data frames to these three stations, or these three stations can concurrently send their respective data frames to the AP. This seems to imply that a WLAN with multi-beam AP can achieve M times the throughput of that with omni-antenna AP. However, IEEE 802.11 [5], the de facto standard for WLANs, employs CSMA/CA (carrier sense multiple access with collision avoidance) mechanism at the MAC

- *Z.-T. Chou and C.-Q. Huang are with the Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung 804, Taiwan. E-mail: ztchou@mail.ee.nsysu.edu.tw; tcihuang@gmail.com.*
- *J. M. Chang is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA. E-mail: morris@iastate.edu.*

Manuscript received 6 May 2012; revised 24 Mar. 2013; accepted 20 June 2013. Date of publication 15 July 2013; date of current version 22 July 2014. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier 10.1109/TMC.2013.85

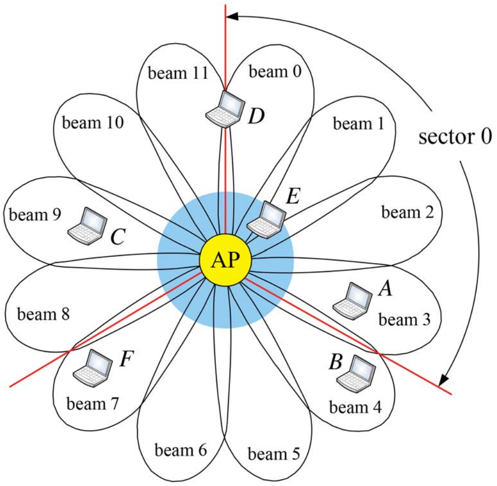


Fig. 1. Example of sectorized multi-beam antenna system. We number the beams and sectors in a clockwise direction. In this example, $M = 3$ and $N = 12$.

layer. Therefore, if directly applying 802.11 to a WLAN with multi-beam AP, we have to face the following the challenges.

1) *Beam-synchronization constraint*. To avoid the co-site interference problem, all sectors at the AP must be in either the transmission mode or the reception mode [19].

2) *Receiver blocking problem*. Referring to Fig. 1, assume that station B in sector 1 intends to send data to the AP when the AP is sending data to station A in sector 0. Since not hearing the directional signal from the AP to station A, station B concludes that the media is free, and then sends data to the AP. On the other hand, due to the beam-synchronization constraint, the AP is unable to receive B's data. Without getting response from the AP, station B may keep sending data until its retry limit is reached, leading to significant bandwidth waste. To make matters worse, since station A is located close to B, B's transmission may corrupt A's reception of data from the AP.

3) *Unnecessary defer problem*. Referring to Fig. 1, assume that station B in sector 1 wants to send data to the AP when the AP is receiving data from station A in sector 0. Clearly, stations A and B can simultaneously transmit their respective data to the AP since they are located in different sectors. However, since A and B are geographically close to each other, station B can hear A's signal and will keep silent according to the rules of CSMA/CA, causing the throughput down.

4) *Beam-overlapping problem*. Due to the imperfection of directional antenna, small portion of beam-overlapping area generally exists for two adjacent beams. Especially, a station located in the sector-overlapping area can hear transmissions from multiple sectors, and multiple sectors at the AP can also hear transmissions from that station. Referring to Fig. 1, assume that stations C and D are simultaneously sending data to the AP. Since both sector 0 and sector 2 can hear the signal from station D, sector 2 will receive collided data from stations C and D.

5) *Back/side-lobe problem*. Even though the AP is equipped with multiple high-gain narrow-beam directional antennas, the negative effects of back/side-lobe problem cannot be totally ignored. Referring to Fig. 1, when station E sends data to the AP, all sectors may receive the signal from E since it is too close to the AP and falls in the back-lobe or side-lobe of many other beams.

6) *Hidden terminal problem*. Referring to Fig. 1, assume that station F wants to send data to the AP when the AP is receiving data from either station B or station E. Since not hearing the signal from either station B or station E, station F infers that the media is free, and then sends data to the AP, which will certainly receive collided data. Note that the necessary conditions to this problem are (i) two stations are out of the range of each other, and (ii) either two stations are in the same sector, or two stations are in different sectors but one of them is too close to the AP.

7) *Multipath rich problem*. In a multipath rich environment, *any* station, say station B, in the coverage of the AP may hear transmissions from all sectors; vice versa, all sectors at the AP can hear transmissions from station B. This implies that in a multipath rich environment, no spatial reuse can be exploited.

8) *Miss-hit problem*. On the basis of DOA (direction of arrival) estimation techniques [14], when a station sends frames to the AP with a smart antenna system, the AP can identify which *beam* (or which *beams*, if beam-overlapping problem, back/side-lobe problem, or multipath rich problem occur) the sending station is located in [19]. However, the beam-location information cached in the AP may be stale and incorrect when mobile stations move. Under such circumstances, the AP may direct a wrong beam for downlink transmission.

1.2 MAC Design Challenges for QoS Provisioning

Wang *et al*. [19] assume that each sector consists of the same number of beams; then on the basis of p-persistent DCF (distributed coordination function), they designed a MAC protocol to carefully address the above-mentioned challenges, excluding the unnecessary defer problem. Tang *et al*. [17] assume that, under the constraint that the total number of beams remains constant, the AP can quickly adjust the sector-configuration so that each sector can consist of different number of beams. Based on this assumption, [17] modified Wang's protocol to additionally mitigate the unnecessary defer problem and unbalanced sector-load problem. Appendix A briefly presents these two protocols [17], [19]. However, we notice that DCF does not provide QoS (*quality-of-service*) mechanisms. This implies that their protocols [17], [19] may not be suitable for real-time multimedia applications. To support QoS, IEEE 802.11e [6] proposes a new coordination function, called HCF (hybrid coordination function), which defines two channel access schemes: EDCA (enhanced distributed channel access) and HCCA (HCF controlled channel access). In Appendix A, we briefly describe the operations of 802.11e. The major advantage of EDCA over DCF in 802.11 is that EDCA supports traffic prioritization. In EDCA, the AIFS (arbitration inter-frame space), $\mathrm{CW}_{\mathrm{min}}$, and, $\mathrm{CW}_{\mathrm{max}}$ of a high-priority frame are respectively smaller than those of a low-priority frame, where $\mathrm{CW}_{\mathrm{min}}$ and $\mathrm{CW}_{\mathrm{max}}$ are the minimum and maximum contention windows (CWs), respectively. By this way, a station with high priority traffic waits, *on average*, less before sending its frame than a station with low priority traffic. The major advantage of HCCA over PCF (point coordination function) in 802.11 is that HCCA enforces that the transmission time of a polled station cannot exceed its TXOP (*transmission opportunity*) limit. This mechanism overcomes the problem of a polled station gaining an inordinate amount of airtime in PCF, which may severely ruin the performance of other admitted stations. However, when applying EDCA and/or HCCA to a WLAN with multi-beam AP, we have to face the following challenges.

1) *Prioritization-induced problem*. The prioritization scheme of EDCA easily induces the unnecessary defer problem and receiver blocking problem. Take Fig. 1 for example. We assume that both stations *A* and *B* want to send data to the AP, which also wants to send data to *A*. Besides, we assume that the priorities of *A* and the AP are equal but higher than that of *B*. If station *A* first wins the contention, *B* will temporarily suppress its transmission, which is unnecessary. If the AP first wins the contention, the data later sent from *B* to the AP will be lost due to the beam-synchronization constraint. Moreover, EDCA may suffer from the *priority reversal problem* [3]: Since the number of random backoff slots is associated with the CW, and the CW is exponentially proportional to the number of retransmission attempts, a high-priority *backlogged* frame may experience a longer waiting time than a low-priority unbacklogged frame. Note that a frame which involved in a collision and must be retransmitted is said to be backlogged [2].

2) *Contention-parallelization problem*. Since multimedia traffic is typically *isochronous* and *time-sensitive*, we hope that real-time stations can promptly seize or reserve the access right. The contention schemes in 802.11 [5] and 802.11e [6] are DCF and EDCA, respectively. Both Wang's protocol [19] and Tang's protocol [17] adopt the *p*-persistent DCF as their contention schemes. From Appendix A, we can know that, in Wang's protocol and Tang's protocol (in DCF and EDCA, respectively), contending stations in the range of each other can concurrently send their RTS (data, respectively) frames to the multi-beam AP only when they *coincidentally* have the *same* backoff time. However, the design purpose of backoff mechanisms is to hope that contending stations can select the *different* number of backoff slots. This implies that backoff-based contention mechanisms [5], [6], [17], [19] are innately hard to be parallelized and thus unable to fully exploit the concurrent transmission/reception capability of the multi-beam AP. Worse yet, due to the nature of *randomness*, backoff-based MAC schemes fail to guarantee the *bounded* contention time.

3) *Power-saving scheduling problem*. Both Wang's protocol and Tang's protocol do not provide power saving mechanisms. MAC protocols can assist mobile stations, which are often powered by batteries, to conserve energy by identifying when they can enter the *doze* state [5]. This implies that how the AP schedules the polling order certainly influences the energy efficiency. Although 802.11e [6] introduces a new power saving mechanism, called S-APSD (scheduled automatic power-save delivery), how the AP schedules the polling order under S-APSD is unspecified. Classical polling-based MAC protocols [15] adopt the *shortest job first* policy to schedule the polling order. However, in Section 2.5, we will show that such an optimal scheduling policy in a WLAN with omni-antenna AP almost becomes the worst scheduling policy in a WLAN with multi-beam AP.

1.3 Objective and Contributions

The objective of this paper is to design an 802.11-compliant MAC protocol that makes full use of the multi-beam AP to provide QoS functionalities while preventing/mitigating all the above-mentioned problems. To achieve our objective, we carefully extend and tailor HCCA and our previously proposed UPCF [3] such that our newly designed protocol, named *M-HCCA* (*multi-beam AP-assisted HCCA*), has the following attractive features.

1) Since M-HCCA is a polling-based scheme, it innately can detect, prevent, mitigate, or resolve all the problems mentioned in Section 1.1 in a simple and effective manner. Since the input of the polling scheduling in M-HCCA includes the beam-location information, the negative effects of unbalanced sector-load problem can be minimized. Moreover, M-HCCA offers a location updating mechanism to promptly renew the beam-location information of a nonresponsive station such that the adverse effects of miss-hit problem can be minimized.

2) M-HCCA adopts the *handshaking* mechanisms, instead of using backoff in the CP (contention period), to accomplish traffic prioritization *during the CFP* (*contention-free period*). Importantly, M-HCCA guarantees that high-priority stations are always admitted to the polling list earlier than low-priority stations.

3) M-HCCA employs the deterministic tree-splitting algorithm as its reservation scheme, which completely rules out random backoff mechanisms to not only boost the contention parallelism but also guarantee the *bounded* reservation time. Especially, during the reservation period of M-HCCA, the AP can adaptively adjust the sector configuration according to the feedback of contending stations to speed up the reservation process.

4) M-HCCA achieves energy conservation via the following three approaches. First, as compared with contention-based MAC protocols, M-HCCA adopts the polling-based access scheme to reduce energy waste on collisions and retransmissions as far as possible. Second, M-HCCA utilizes the PL (polling-list) frame to let stations which cannot partake in the polling activities immediately return to the doze state. Last, M-HCCA adopts the energy-conserving scheduling such that stations which should partake in the polling activities can spend as little awake time as possible.

5) M-HCCA adopts the cross-layer rate adaptation scheme to regulate the audio/video source rate such that the demanded airtime of each admitted station can never exceed its TXOP limit. A valuable by-product of such scheme is that it can not only avoid the performance anomaly phenomenon [20] but also simplify the design of admission control scheme.

6) Since the length of the maximum CFP duration is limited, we integrate the *run-time admission control* mechanism into the reservation procedure such that, even in a multipath environment, the AP can admit as many

newly real-time stations as possible while maintaining QoS guarantees made to already-admitted stations.

7) We consider the backward compatibility in the design of M-HCCA. Since only operating in the CFP, M-HCCA can coexist with 802.11 DCF and 802.11e EDCA.

2 The M-HCCA Protocol

2.1 Models and Assumptions

Depending on the antenna system's capability, we consider two types of multi-beam APs: the *fixed* multi-beam AP [19] and the *reconfigurable* multi-beam AP [17]. As for the fixed multi-beam AP, the set of beams in sector S_i is always $\{b_{i\times\omega}, \dots, b_{(i+1)\times\omega-1}\}$, for all $0 \leq i \leq M-1$; for convenience, we denote $S_i = \{b_{i\times\omega}, \dots, b_{(i+1)\times\omega-1}\}$. On the other hand, under the constraint that the total number of beams remains constant, the reconfigurable multi-beam AP can adjust the sector-configuration in a short period of time such that each sector may consist of different number of beams [17].

To enjoy the WiFi services, a mobile station should first discover the presence of APs by *passive scanning* or *active scanning* [5]. In passive scanning, a mobile station should keep silent until receiving beacons from the APs. In active scanning, a mobile station needs to first wait for *ProbeDelay* and then broadcast a probe request to solicit responses from the APs. After the receipt of beacons or probe responses, that station then attempts to associate or reassociate with a particular AP. When the (re)association request is granted, the AP responds with a status code of 0 (successful) and the AID (*association identifier*). The AID is an integer used to logically identify the mobile station. The AP can thus maintain a list of finite stations associated within its BSS and updates it whenever a new station joins or a station leaves the BSS. Due to security considerations, in M-HCCA, a station with real-time traffic can join the polling list only after (re)association. Especially, M-HCCA disables the *CF-Pollable* and *CF-Poll Request* subfields of the *capacity information* field in (re)association request frames [5]. Instead, M-HCCA offers a new reservation mechanism to let real-time stations quickly get on/off the polling list without relying on the reassociation.

2.2 CFP Structure and Timing

In a WLAN cell, known as the *basic service set* (*BSS*), the AP takes charge of airtime allocation and makes two coordination functions, DCF and M-HCCA, alternative, with a CFP (during which M-HCCA is active) followed by a CP (during which DCF is active), which are together referred as a *superframe*. The AP normally operates in the multi-beam antenna mode during the CFP, except in a multipath rich environment. Referring to Fig. 2, at the nominal start of each CFP, known as the TBTT (*target beacon transmission time*), every station shall wake up and remain awake to listen for the PL (*polling list*) frame; meanwhile, the AP continuously monitors the channel and then seizes its control by broadcasting the *beacon* frames after the *PIFS* medium idle time. In M-HCCA, as shown in Fig. 2, the CFP is divided into three periods: the *prioritization period*, the *collision resolution period*, and the *polling period*. The first two periods are together called the *reservation period*. During the prioritization period, the AP performs a series of handshakes to ensure that high-priority stations are always admitted to the polling list earlier than low-priority stations. During the collision resolution period, the AP performs a deterministic tree-splitting algorithm to probe which stations undergo the prioritization period desire to join the polling list. Once the reservation process terminates, the AP broadcasts the PL frames to announce the start of the polling period. Upon examining the PL frame, a station that can be neither a sender nor a receiver during the polling period may return to the doze state. Note that if being equipped with the reconfigurable multi-beam antennas, the AP can adaptively adjust the sector configuration during the collision resolution period and the polling period to speed up the reservation process and minimize the average awake time of all polled stations, respectively. After the close of the polling period, the AP broadcasts the *CF-End* frames to let all stations enter the CP. During the CP, the AP runs DCF and operates in the omni-antenna mode. Thus 802.11-compliant stations that do not implement M-HCCA can still communicate with the AP during the CP.

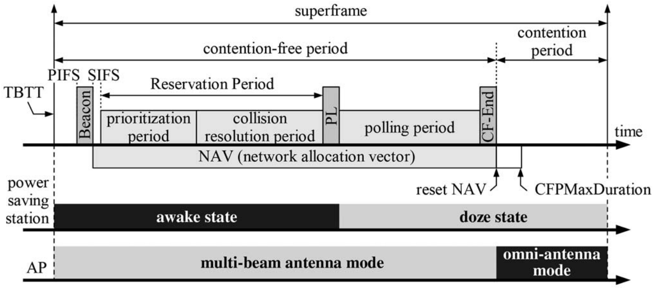


Fig. 2. Proposed superframe structure for a WLAN with multi-beam AP.

Clearly, the maximum length of CFP, denoted by *CFPMaxDuration*, shall be limited to allow coexistence between DCF and M-HCCA traffic. As per 802.11 [5], the minimum length of CP, denoted by CP$_{min}$, is the time needed to transmit and acknowledge one maximum-sized MPDU (*MAC protocol data unit*); namely, CP$_{min}$ = DIFS + SIFS + $(L_{max\text{MPDU}} + L_{\text{ACK}})/R_{\min}$, where L_{ACK} is the length of ACK frame and $R_{\min}$ is the minimum PHY rate. Thus we have CFPMaxDuration = SF − CP$_{min}$, where SF is the superframe length. Since the length of CFPMaxDuration is limited, the overrun of the reservation process may shorten the polling period, violating the quality of already-admitted connections. Hence a run-time admission control is established to assist the AP in determining when the reservation period shall be terminated. In particular, when the polling list size reaches the saturation point (see Section 2.6), the AP may directly dive into the polling period at the start of CFP without first performing the reservation procedure.

2.3 Prioritization Procedure

The purposes of the prioritization procedure are to provide multiple levels of priorities and to ensure the freedom from the priority reversal problem. In M-HCCA, priority levels (known as *access categories* [6]) are numbered from 0 to H, with H denoting the highest priority level. A frame with priority 0 (i.e. best-effort traffic) should be sent via the DCF. On the other hand, only the active real-time station that has a stream with priority level ranging from 1 to H can

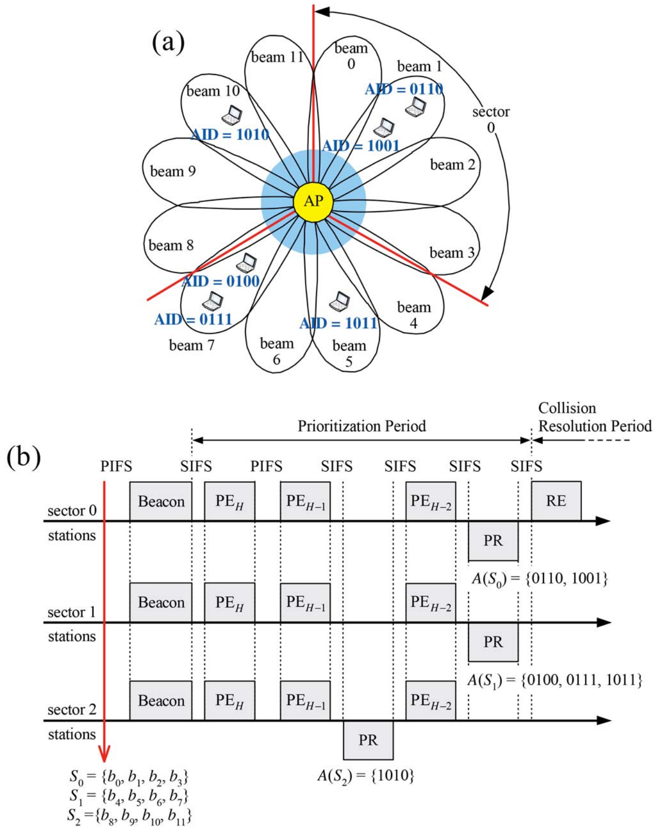


Fig. 3. Example of the prioritization procedure. (a) The initial WLAN configuration (b) $A(S_i)$ denotes the set of stations (AIDs) in sector S_i responding to the enquiry frame.

participate in the reservation process. Note that a real-time station is called *active* if it desires to get on the polling list. Besides, a *stream* is a continuous sequence of frames that have the same source, destination, and access category.

From the start of CFP to the end of the prioritization period, the reconfigurable multi-beam AP adjusts the sector configuration such that $S_i = \{b_{i\times\omega}, \ldots, b_{(i+1)\times\omega-1}\}$, for all $0 \le i \le M-1$. The reason is that the AP does not know the location distribution of active stations; if they are uniformly distributed in the BSS, such a configuration can let the AP discover the maximum number of sectors containing active stations.

As shown in Fig. 3(b), after broadcasting a beacon frame and waiting for SIFS, each sector at the AP sends the PE$_H$ (*priority enquiry*) frame to invite every active station whose priority level equals H to reply with the PR (*priority response*) frame. On receiving the PE$_H$ frame, an active station with priority level H shall acknowledge a PR frame after a SIFS period. At the end of the handshake, each sector at the AP obtains the ternary feedback information according to stations' responses: (i) IDLE: The sector does not receive any PR frames. (ii) SINGLE: The sector successfully receives a single PR frame. In this case, the sector will place the vector (*AID*, *beam-location*) of that active station on the polling list. (iii) COLLISION: This event occurs if the sector encounters neither IDLE nor SINGLE.

If the conclusions of the current handshakes are that at least one sector encounters a SINGLE event and no other sectors encounter COLLISION events (or all sectors encounter IDLE events, respectively), the AP will proceed to the next handshakes by issuing the PE$_{H-1}$ frames after an elapsed SIFS (PIFS, respectively). This priority probing process keeps running until the delivery of the PE$_1$ frames, the occurrence of a COLLISION event, or a failure in the run-time admission test (see Section 2.6), whichever comes first. Especially, once at least one sector perceives a COLLISION event, the AP immediately sends RE (*registration enquiry*) frames to announce the start of the collision resolution period.

Fig. 3(b) illustrates how the prioritization procedure works. In this example, we assume that there are 15 associated stations in the BSS. Fig. 3(a) shows that stations 4, 6, 7, 9, 10, and 11 intend to join the polling list. In the first round, all sectors send the PE$_H$ frame and no one responds. In the second round, only station 10 replies with the PR frame and thus successfully joins the polling list. At the end of the third round, sectors 0 and 1 encounter COLLISION events, and then the AP starts the collision resolution procedure.

2.4 Collision Resolution Procedure

The purpose of the collision resolution procedure in M-HCCA is for the AP to discover which active stations bring the COLLISION events at the end of the the prioritization period. Theoretically, a multi-access algorithm that is suitable in the collision resolution period had better satisfy three properties: simplicity, *parallelizability*, and *bounded* collision resolution period. According to these criteria, the *identifier*-based tree-splitting algorithm [2] is an appropriate choice. Before presenting our collision resolution procedure, we need to define some notations. We assume that there are n stations associated with the AP and each station is assigned a unique AID $a \in \mathcal{A} = \{1, 2, \ldots, n\}$, where $n \le 2007$ [5]. The AID a can be represented by a binary k-tuples $(a_k a_{k-1} \cdots a_2 a_1)$, where $a_i \in \{0, 1\}$ and $k = \lceil \log_2 n \rceil$. Note that the i-th bit corresponds to the i-th *dimension*. For example, let $\mathcal{A} = \{1, 2, 3\} = \{01, 10, 11\}$. We can partition the set $\mathcal{A}$ along the first dimension into two subsets $\{*0\} = \{10\}$ and $\{*1\} = \{01, 11\}$, where "$*$" means "*don't care*." Given a set $\mathcal{A}$ of binary strings, the set $\mathcal{A}\otimes(dim, value)$ is defined by letting all the *dim*-th bit values of the strings in $\mathcal{A}$ be equal to *value*, where $1 \le dim \le k$ and $value \in \{0, 1\}$. For example, let $\mathcal{A} = \{10{*}0\}$. Then we have $\mathcal{A}\otimes(3, 1) = \{11{*}0\} = \{1100, 1110\}$ and $\mathcal{A} \otimes (3, *) = \{1{*}{*}0\} = \{1000, 1010, 1100, 1110\}$.

The basic idea of the tree-splitting algorithm is to use the stack to implement a *preorder traversal* of the *dimension splitting tree*. Specifically, when COLLISION events occur, the AP splits the set $\mathcal{A}$ of stations involved in collisions into two subsets, $\mathcal{A}_1$ and $\mathcal{A}_2$, along a dimension *dim*. The AP first recursively resolves the collisions of $\mathcal{A}_1$, and then resolves the collisions of $\mathcal{A}_2$ independently. Besides the *address* partition, the reconfigurable multi-beam AP can use the *beam* partition mechanism to speed up the collision resolution process. Fig. 4 presents the tree-splitting algorithm. We assume that the close of the prioritization period results from the transmission of multiple PR$_h$ frames in at least one sector, where $1 \le h \le H$. During the collision resolution period, the AP first popes a vector (*dim*, *value*, *SC*) from its local stack, and then updates the sector configuration *SC* and the set of binary strings, *AddressPattern* $\mathcal{A}$, according to the popped vector. Next, each sector sends the RE frame which contains the value

```
**Sector_Configuration_Adjustment**($SC$: sector configuration)
01  **if** (the AP is equipped with *fixed* multi-beam antennas)
02      **return** $SC = \{S_i = \{b_{i\times\omega}, \cdots, b_{(i+1)\times\omega-1}\} \,|\, 0 \le i < M\}$;
03  $C = \{ B_i \mid$ all beams in the set $B_i$ perceive the COLLISION
        events and the indices of these beams are consecutive $\}$;
  // Note that $C$ is a collect of sets.
04  $B^* = \arg\max_{B_i \in C}\{|B_i|\}$;
  // $|B_i|$ denotes the number of beams in the set $B_i$.
05  **if** ( $M-1 \le |B^*| \le (M-1)\omega$ ) {
06      *evenly* partition the set of beams in $B^*$ into $M-1$
          subsets $B_1^*, B_2^*, \cdots, B_{M-1}^*$;
07      $S_0 = \{b_0, \cdots, b_{N-1}\} \setminus B^*$;   $S_i = B_i^*, \forall\, 1 \le i \le M-1$;
08      **return** $SC = \{S_0, S_1, \cdots, S_{M-1}\}$; }
09  **else**
10      **return** $SC$;

**Collision Resolution Procedure**
11  Let $\{d_1, d_2, \cdots, d_k\}$ be the random permutation of
    $\{1, 2, \cdots, k\}$, where $k = \lceil \log_2 n \rceil$.
12  $\mathcal{A} = \{1, 2, \cdots, n\}$;
    // Initially, $\mathcal{A}$ contains all associated stations.
13  STACK $= \varnothing$;   // The AP maintains a local stack.
14  $SC$ = Sector_Configuration_Adjustment($SC$);
15  PUSH$(1, 0, SC)$;
    // The AP pushes the vector $(1, 0, SC)$ onto the stack.
16  **while** (STACK $\neq \varnothing$) {
17      $(dim, value, SC)$ = POP();
    // the AP popes a vector from its stack.
18      $\mathcal{A} = \mathcal{A} \otimes (d_{dim}, vlaue)$;
    // the AP updates the AddressPattern $\mathcal{A}$.
19      **for** $(i = dim + 1;\ i \le k;\ i{+}{+})$
20          $\mathcal{A} = \mathcal{A} \otimes (d_i, *)$;
    // This for-loop controls the level of the splitting tree.
21      all sectors in $SC$ **send** RE$(h, \mathcal{A})$;
22      $status$ = **receive**(RR(AID), *beam-location*);
    /* On receiving the RE$(h, \mathcal{A})$ frame, the active station
    with priority $h$ and AID $\in \mathcal{A}$ shall reply with an RR frame
    including its AID. The AP updates the channel state
    variable $status$ according to received RR frames. */
23      **switch** ($status$) {
24          **case** ((at least one sector encounters a SINGLE event)
                **and** (no sectors encounter COLLISION events)):
25              the sectors encountering SINGLE events place the
                  vectors of (*AID*, *beam-location*) on the polling list;
26              **if** ($value == 0$)   PUSH($dim$, 1, $SC$);   **break**;
27          **case** (all sectors encounter IDLE events):
28              **if** ($value == 0$)   PUSH($dim$, 1, $SC$);   **break**;
29          **case** (at least one sector meets a COLLISION event):
30              the sectors encountering SINGLE events place the
                  vectors of (*AID*, *beam-location*) on the polling list;
31              $SC$ = Sector_Configuration_Adjustment($SC$);
32              **if** ($value == 0$)
33                  PUSH($dim$, 1, $SC$);
34              PUSH($dim + 1$, 0, $SC$);   **break**;
            // To explore the next level subtree.
35      }  //  end of **switch**
36  }  // end of **while**
```

Fig. 4. Collision resolution procedure executed by the AP.

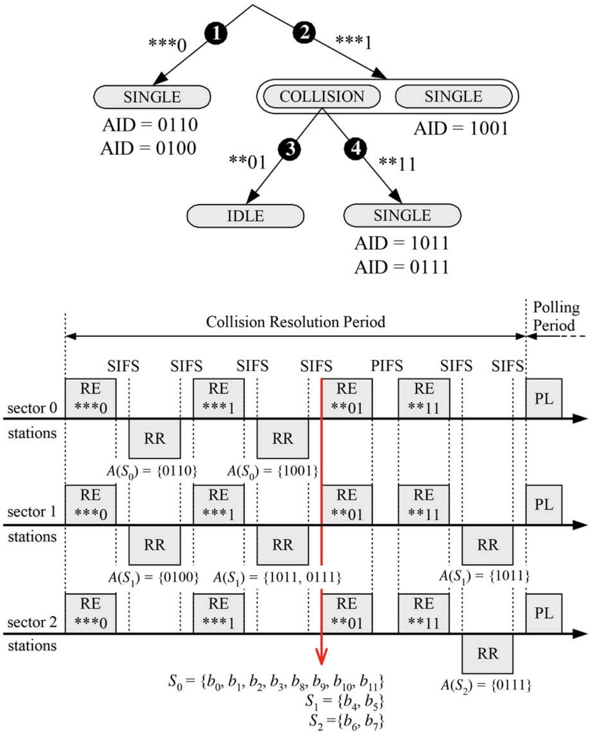


Fig. 5. Example of the collision resolution procedure. The tree structure represents a particular pattern of IDLEs, SINGLEs, and COLLISIONs resulting from a sequence of address partition and beam partition.

of h and the AddressPattern $\mathcal{A}$ to invite active stations to reply with the RR (*registration response*) frames. Upon receipt of the RE$(h, \mathcal{A})$ frame, the active station with priority level h and AID $\in \mathcal{A}$ shall acknowledge an RR frame. At the end of the handshakes, the AP pushes the proper vector(s) onto its local stack according to stations' responses (SINGLE/IDLE/COLLISION). If a sector successfully receives a single RR frame including sender's AID, then that sector adds the vector (*AID*, *beam-location*) to the polling list. In particular, when the COLLISION events occur, the reconfigurable multi-beam AP can further adjust the sector configuration (by calling the function Sector_Configuration_Adjustment(*SC*)) such that the areas suffering from severer collisions could be covered by more number of sectors. Since each sector is equipped with a transceiver, this method may increase the number of to-be-discovered active stations in the next handshakes. This AID probing process will keep running until the emptiness of the stack or a failure in run-time admission test (see Section 2.6), whichever comes first.

Continuing the example of Fig. 3, Fig. 5 illustrates how the collision resolution procedure works. In the first round, each sector sends the RE frame with $\mathcal{A} = \{{*}{*}{*}0\}$, asking for responses. Since stations 4 and 6 reply with the RR frames, the AP adds vectors (*AID*, *beam-location*) $= (0100, b_7)$ and (*AID*, *beam-location*) $= (0110, b_1)$ to the polling list. In the second round, the AP probes the set $\mathcal{A} = \{{*}{*}{*}1\}$, and only sector 1 encounters the COLLISION event. Then the reconfigurable multi-beam AP not only halves the range of $\mathcal{A}$ (i.e. $\mathcal{A} = \{{*}{*}01\}$) but also adjusts the sector configuration (i.e. $S_0 = \{b_0, b_1, b_2, b_3, b_8, b_9, b_{10}, b_{11}\}$, $S_1 = \{b_4, b_5\}$, and $S_2 = \{b_6, b_7\}$). This time, all sectors encounter IDLE events. Hence the AP can skip over large chunks of the address space (i.e. $\{{*}{*}01\}$) that have no active stations. Finally, thanks to the sector reconfigurability, the AP can simultaneously discover two stations 7 and 11 at the end of the forth round. Now, the stack becomes empty and then each sector sends the PL (polling list) frame to let each active station know whether it has been successfully placed on the polling list. To ensure *fairness* with the tree-splitting algorithm, the sequence of dimensions the AP explores shall be randomized in each CFP. (See Fig. 4, line 11.) Essentially, the tree-splitting operation is that of polling, with the AP adaptively controlling the sector configuration and the number of allowably contending stations to finally

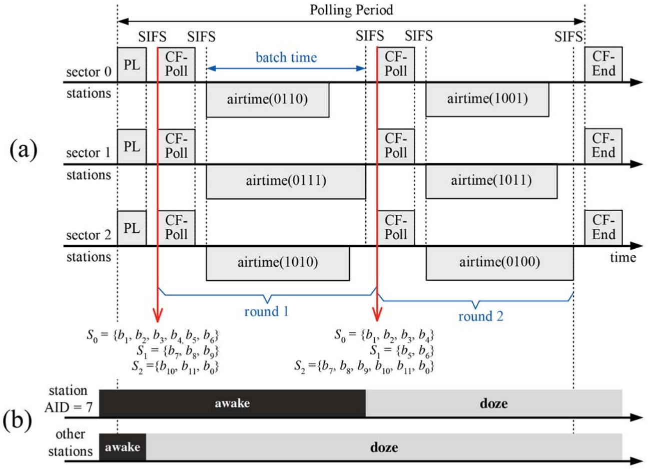


Fig. 6. (a) Shows an example of the polling procedure in a WLAN with reconfigurable multi-beam AP. (b) Shows the power management operation.

identify each active station. The average overhead of the collision resolution operation is expected to be relatively low since the tree-splitting algorithm rules out random backoff and is hence easier to be parallelized than backoff-based contention schemes [17], [19]. In Appendix B, we will quantitatively examine and confirm these issues.

2.5 Polling Procedure and Energy-Conserving Scheduling

In M-HCCA, at the start of the polling period, the AP broadcasts the PL frames to announce which stations shall partake in the polling activities. Note that the PL frame may include the AIDs of the stations to whom the AP intends to send the real-time data. On inspecting the PL frame, a station that can be neither a sender nor a receiver during the polling period may enter the doze state. Fig. 6(a) illustrates how the polling procedure works. In Fig. 6(a), the AP concurrently polls stations 6, 7, and 10 in the first round, and then concurrently polls stations 4, 9, and 11 in the second round. As depicted in Fig. 6(a), the reconfigurable multi-beam AP can adjust the sector configuration before sending the CF-Poll frames. Note that, from Fig. 6(a), we can see that due to the beam-synchronization constraint, sectors 0 and 2 cannot proceed to the second round before station 7 finishes its uplink transmission.

On the other hand, in M-HCCA, each admitted station is polled *exactly once* during the entire polling period. During the reservation period, an active real-time station, say A_i, shall use the PR/RR frame to inform the AP its demanded airtime in the current CFP. In case station A_i is admitted, it shall piggyback $airtime(A_i)$ with the data frame to declare its demanded airtime in the *next* polling period. To ensure airtime fairness and avoid the performance anomaly [20], in M-HCCA, the beacon frame specifies the *limit* of TXOP for each access category, TXOP[AC]. In what follows, we show how M-HCCA fulfils the goal that the demanded airtime of each admitted station A_i, $airtime(A_i)$, with access category h can never exceed $\mathrm{TXOP}_{A_i}[h] = \mathrm{TXOP}[h]$.

During the CFP, the AP can measure the uplink channel quality between the AP and the admitted station A_i in terms of SNR (signal-to-noise ratio) when station A_i sends frames to the AP. Then the AP can apply the existing SNR-based PHY rate adaptation scheme [7] to determine the highest allowable PHY rate R^* for station A_i such that the ratio of *data* frame loss due to channel errors could be no more than a predefined threshold, say 3%. On the other hand, the current scalable video/audio codec technologies (such as scalable audio coder G.729.1 [18] and scalable video coding (SVC) scheme [13]) can enable the sender station to adjust the source bit rate *on the fly* according to the available bandwidth such that the best possible streaming quality can be achieved in time [13], [18]. Thus when the AP polls station A_i to ask it to use the PHY rate R^* to upload data frames to the AP, station A_i first employs (1) to estimate the upper bound of its individual instantaneous throughput $\mathcal{G}_{A_i}$.

$$\mathcal{G}_{A_i} \approx \frac{\mathrm{TXOP}_{A_i}[h] \times R^* \times \left(1 - \overline{\mathrm{FER}}\right)}{\mathrm{SF}}, \qquad (1)$$

where SF is the superframe length and $\overline{\mathrm{FER}}$ is the observed frame error rate at the AP in a fixed interval (e.g., every 100 ms). Let V be the set of allowed bit rates of the layered audio/video stream. Station A_i can hence adjust the source audio/video bit rate r according to the following formula.

$$r = \max\left\{v \mid v \in V \text{ and } v \le \mathcal{G}_{A_i}\right\}. \qquad (2)$$

Via the above cross-layer rate adaptation scheme, M-HCCA guarantees that each admitted station A_i with access category h requires airtime only $airtime(A_i) = (r \times \mathrm{SF})/R^* \le (\mathcal{G}_{A_i} \times \mathrm{SF})/R^* \le \mathrm{TXOP}_{A_i}[h] = \mathrm{TXOP}[h]$ in a superframe.

To conserve energy, an admitted station may remain awake for *only* a portion of the polling period through the time that the station finishes sending or receiving data frames. Fig. 6(b) shows the power management operation of M-HCCA. From Fig. 6, we can observe that since the transmission time of each polled station may be different, how the AP schedules the polling order can strongly influence the energy efficiency of M-HCCA. Therefore, we want to design an energy-efficient scheduling algorithm that meets the following two objectives.

O1. The length of the polling period should be as short as possible. Since admitted stations may sleep during the CP, the shorter the polling period (thus the longer the CP), the better.

O2. During the polling period, the average awake time of admitted stations should be as short as possible.

Traditionally, polling-based MAC protocols [3], [15] adopt the *shortest job first* (or called *shortest station-airtime first*) policy to schedule the polling order. When applying this policy to a WLAN with *fixed* multi-beam AP, in a round, each sector at the AP will first select the yet-to-be-scheduled station that currently has the *shortest* demanded airtime. Take Fig. 3(a) as an example. We assume that the demanded airtimes of stations 4, 6, 7, 9, 10, and 11 are 360 μs, 300 μs, 400 μs, 300 μs, 350 μs, and 320 μs, respectively. When employing this policy, the AP will poll stations 6, 10, 11 in the first round, poll stations 4, 9 in the second round, and poll station 7 in the third round. Let us call the *set of polled stations in a round* as a *batch* and the *maximum demanded airtime of a polled station in a round* as the *batch time*. The shortest station-airtime first policy yields the total batch time of 1110 μs. In fact, if we regard the *AP* and the *demanded airtime* of an admitted station as the *machine* and *job size*, respectively, the scheduling problem that aims to achieve **O1** is similar to the minimum makespan scheduling problem on a batch processing machine [10]. By pairwise job interchange argument [10, pp. 36–37], we can easily prove that in a WLAN with fixed multi-beam AP, the polling period will reach the minimum length if the AP adopts the *largest job first* (or called *largest station-airtime first*) scheduling policy. Continuing the same example, the largest station-airtime first policy will yield the total batch time of 1080 μs.

Next, we consider the polling scheduling in a WLAN with *reconfigurable* multi-beam AP. If we still apply the largest station-airtime first policy (i.e., in a round, the AP first selects the yet-to-be-scheduled station whose demanded airtime is largest *from all beams*) to the above example, it will require three rounds and yield the total batch time of 1060 μs, which is far from the optimal scheduling result shown in Fig. 6(a). Let us more formally describe the scheduling problem. Given n admitted stations $A_1, \ldots, A_n$, let the parameter $p_{i,k}$ be 1 if station A_i is located in beam b_k and 0 otherwise. Besides, let $x_{i,j}$ denote a decision variable that assumes the value 1 if station A_i is polled in round j and 0 otherwise. The scheduling problem aiming to minimize the polling period length in a WLAN with reconfigurable AP can be modeled as the following multi-objective optimization problem.

$$\min z_1 = \sum_{j=1}^{n} \max_{1 \le i \le n} \{x_{i,j} \times \mathit{airtime}(A_i)\}$$

$$\min z_2 = \sum_{j=1}^{n} \bigvee_{i=1}^{n} x_{i,j}$$

$$\text{subject to } \sum_{j=1}^{n} x_{i,j} = 1, \sum_{i=1}^{n} x_{i,j} \le M, \sum_{i=1}^{n} p_{i,k} x_{i,j} \le 1.$$

Since the multi-objective integer programming problems are NP-hard in general [10], we propose a *new* variant, named the *largest beam-airtime first* scheduling policy, to efficiently find out the near-optimal solution. Specifically, in a round, the AP first selects the admitted station whose demanded airtime is largest from the beam that currently has the largest beam-airtime, where the *beam-airtime* is defined as the sum of demanded airtimes of yet-to-be-scheduled stations in a beam. Obviously, this scheduling policy tries to achieve only the objective **O1**. Therefore, we need to additionally call for the *shortest batch first* scheduling policy, which provides the minimum average waiting (awake) time for a set of batches [10]. In other words, M-HCCA adopts *two-phase* scheduling algorithm such that objectives **O1** and **O2** can be both satisfied. Specifically, in the first phase, the AP adopts the largest station-airtime first or the largest beam-airtime first to select each batch, namely, the set of to-be-polled stations in a round. In the second phase, the AP employs the shortest batch first to arrange the batch order. Appendix C presents the detailed scheduling algorithm whose running time is $O(n \log n)$. Continuing the above-mentioned example, Fig. 6(a) shows the result of our two-phase scheduling algorithm. Note that if we apply the shortest station-airtime first policy to the same example, it will require three rounds and yield the total batch time of 1110 μs, which is the worst scheduling result.

2.6 Run-Time Admission Control

Since the length of CFPMaxDuration is limited, the purpose of run-time admission control is for the AP to determine when to close the reservation process in order not to violate the airtime assurances made to already admitted stations. Existing admission control mechanisms [6] often require that the mobile station should submit its QoS requirements when making a reservation, and then the AP performs the admission test to decide whether to accept/reject that connection request according to the available resources. However, such a traditional approach is not suitable for M-HCCA in that the reservation request/response frame exchange failing the admission test simply wastes the scarce radio bandwidth. Instead, during the reservation period, M-HCCA adopts the *mobile-assisted* admission control scheme: Before sending the PE/RE frames, the AP first evaluates the airtime usage based on the demanded airtimes of admitted stations. If the execution of the PE/PR or RE/RR handshakes will cause the violation of airtime assurances made to admitted stations, the AP directly dives into the polling period; otherwise, the AP sends the PE/RE frames which piggyback the information about the *remaining available airtime* (RAAT). Upon reception of the PE/RE frame, active real-time stations take the admission test and check whether the RAAT is sufficient to meet their QoS requirements. Those who pass the admission test can reply with the PR/RR frames and report their QoS needs; while those who fail the admission test shall abort the contention in the remaining reservation period and wait for the next CFP. A valuable by-product of this approach is that the contending traffic load may be further reduced, making the tree-splitting algorithm more efficient. Importantly, the following two principles guide the design of run-time admission control algorithm.

P1. The AP must guarantee that the progress of the reservation process will not affect the reserved demanded airtime *airime*(A_i) of each admitted station A_i on the polling list $\mathcal{L}$.

P2. Referring to Fig. 7, it is possible for contention-based service to run past the nominal start of the CFP, i.e. TBTT. As per 802.11 [5], in the case of a busy medium due to DCF traffic, the CFP is *foreshortened* and the beacon shall be delayed for the time needed to complete the existing DCF frame exchange. Such a phenomenon is called *stretching* and the length of the stretching time T_s may be up to $\widehat{T}_s = (L_{\text{RTS}} + L_{\text{CTS}} + L_{max\text{MPDU}} + L_{\text{ACK}})/R_{\min} + 3 \times \text{SIFS}$. The AP must make sure that the *upper bound of the demanded airtime* $\text{TXOP}_{A_i}[h]$ of each admitted station A_i can be guaranteed during the entire *stream lifetime* even in the worst case scenario, that is, $T_s = \widehat{T}_s$, $\mathit{airtime}(A_i) = \text{TXOP}_{A_i}[h]$ for all $A_i \in \mathcal{L}$, and the AP equivalently runs in the omni-antenna mode during the entire CFP because the multipath rich problem occurs or all stations move in the all-beam-overlapping area (refer to Fig. 9).

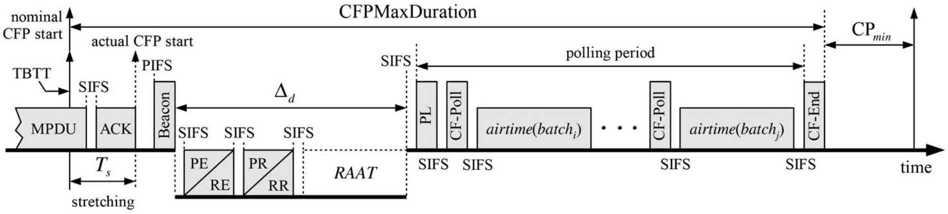


Fig. 7. Run-time admission control process and the timing relationship between *RAAT* and Δ_d.

We now introduce some notations used to facilitate the presentation of the admission control algorithm shown in Fig. 8.

- Let O_{CFP} denote the fixed overhead in a CFP. If $\mathcal{L} \neq \emptyset$, we have $O_{\text{CFP}} = \text{PIFS} + T_{\text{beacon}} + T_{\text{PL}} + T_{\text{CF-End}} + 2 \times \text{SIFS}$.
- During the reservation period, we let

$$\delta_1 = \begin{cases} T_{\text{PE}} \text{ if the AP sends out the PE frame,} \\ T_{\text{RE}} \text{ if the AP sends out the RE frame.} \end{cases}$$

$$\delta_2 = \begin{cases} T_{\text{PR}} \text{ if a mobile replies the PR frame,} \\ T_{\text{RR}} \text{ if a mobile replies the RR frame.} \end{cases}$$

- Before the start of the reservation period, the AP first executes the polling scheduling algorithm shown in Appendix C. Referring to Fig. 7, let $batch_i = \{A_{i_1}, \ldots, A_{i_k}\}$ be the set of to-be-polled stations in a round, where $k \leq M$. Let $airtime(batch_i) = \max_{1 \leq j \leq k}\{airtime(A_{i_j})\}$. Then we define two auxiliary variables Δ_d and Δ_g to assist the AP in verifying whether **P1** and **P2** are respectively satisfied, where

$$\Delta_d = \text{CFPMaxDuration} - \Bigg[T_s + O_{\text{CFP}} + \sum_{\bigcup batch_i = \mathcal{L}} (T_{\text{CF-Poll}} + airtime(batch_i) + 2\text{SIFS})\Bigg] \quad (3)$$

and

$$\Delta_g = \text{CFPMaxDuration} - \Bigg[\widehat{T}_s + O_{\text{CFP}} + \sum_{A_i \in \mathcal{L}} \big(T_{\text{CF-Poll}} + \text{TXOP}_{A_i}[h] + 2\text{SIFS}\big)\Bigg]. \quad (4)$$

Referring to Fig. 7, Fig. 8 presents the admission control operations performed cooperatively by the AP and all active real-time stations during the reservation period.

```
After sending the beacon frames, the AP computes $\Delta_d$, $\Delta_g$,
 and the variable $RAAT = \Delta_d - (\delta_1 + \delta_2 + 3 \times \text{SIFS})$;
 /* $RAAT$ denotes the remaining available airtime if the
  AP proceeds to the next PE/PR or RE/RR handshakes. */
while ($\Delta_g \geq \min_{AC \in \{\text{voice, video}\}}\{\text{TXOP}[AC]\}$ and
   $RAAT > 0$ and (reservation process is not finished)) {
  each sector sends the PE/RE frame which announces
   (AC, $\Delta_g$, $RAAT$);   /* On receiving the PE/RE frame,
   each active real-time station, say $A_i$, with access
   category AC takes the following admission test. */
  if ($\text{TXOP}_{A_i}[AC] \leq \Delta_g$ and $airtime(A_i) \leq RAAT$)
      station $A_i$ replies with the PR/RR frame and declares
       its demanded $airtime(A_i)$ in the current CFP;
  $status$ = receive(PR or RR);
 /* The AP updates the channel state variable $status$
   according to the received PR/RR frames. */
  switch ($status$) {
    case (at least one sector encounters the SINGLE event):
 // Assume that the AP correctly receives $k$ PR/RR frames.
      if ($\Delta_g < k \times (T_{\text{CF-Poll}} + \text{TXOP}[AC] + 2 \times \text{SIFS})$) {
        $j = \lfloor \Delta_g / (T_{\text{CF-Poll}} + \text{TXOP}[AC] + 2 \times \text{SIFS}) \rfloor$;
        the AP randomly admits $j$ contending stations; }
      else
        the AP admits all these $k$ contending stations;
      AP records (AID, beam-location) of admitted stations;
      $\Delta_g = \Delta_g - \frac{\Delta PL}{R_{\min}} - \sum_{A_i \in \text{SINGLE}} (\text{TXOP}_{A_i}[AC]$
        $+T_{\text{CF-Poll}} + 2\text{SIFS})$;
      $\Delta_d = \Delta_d - (\delta_1 + \delta_2 + \max_{A_i \in \text{SINGLE}}\{airtime(A_i)\}$
        $+T_{\text{CF-Poll}} + \frac{\Delta PL}{R_{\min}} + 4 \times \text{SIFS})$; break;
 /* $\Delta PL$ denotes the increased size of the PL frame.
 "$A_i \in \text{SINGLE}$" means that $A_i$ is now admitted by the AP. */
    case (all sectors encounter the IDLE events):
      $\Delta_d = \Delta_d - (\delta_1 + \text{PIFS})$; break;
    case (at least one sector meets the COLLISION event):
      $\Delta_d = \Delta_d - (\delta_1 + \delta_2 + 2 \times \text{SIFS})$; break;
  } // end of switch
  $RAAT = \Delta_d - (\delta_1 + \delta_2 + 3 \times \text{SIFS})$;
} // end of while
```

Fig. 8. Admission control algorithm of M-HCCA.

3 Resolution of Corner Cases

3.1 Contention Related Corner Cases

The contention related problems and their respective prevention/resolution methods in M-HCCA are listed as follows.

1) *Beam-synchronization constraint, receiver blocking problem, unnecessary defer problem, and hidden terminal problem.* During the CFP, mobile stations can send frames only when they are allowed to do so by the AP. Hence the beam-synchronization constraint can be naturally satisfied; besides, the receiver blocking problem, unnecessary defer problem, and hidden terminal problem induced by carrier sensing will never occur. During the CP, the AP operates in the omni-antenna mode. Hence the receiver blocking problem and unnecessary defer problem induced by directional signals will never occur; besides, the hidden terminal problem can be effectively alleviated by the handshake of RTS/CTS frames [4].

2) *Starvation.* When several real-stations contend to join the polling list, lower-priority stations will be blocked if they have no chance to send out PR frames during the entire prioritization period. We could adopt the *aging* policy (As time progresses, so does the priority of the streaming.) to conquer the problem of starvation.

3) *DCF-based interference.* To ensure the correctness of M-HCCA, we must lock out the DCF-based access during the CFP. If all stations are in the range of each other,

the transmissions of M-HCCA during the CFP are separated only by *SIFS* or *PIFS*. Under such circumstances, the AP can naturally safeguard its control of the medium against the DCF-based interference even if the beacon frame is lost. However, in a WLAN, some stations may be out of range of other stations. Thus 802.11 further employs the *virtual* carrier sensing to prevent the DCF-based interference. First, all control frames sent during the CFP set the NAV (network allocation vector) to CFPDurRemaining (i.e., remaining time of the current CFP). More importantly, in 802.11, both beacon and probe response frames include timestamp, CFPPeriod (i.e., superframe length), CFPMaxDuration, and CFPDurRemaining. Once a station has received a beacon or probe response, it can infer the TBTT of every superframe according to these parameters. Since CFPMaxDuration is a *constant*, each station can *preset* its NAV to the CFPMaxDuration *by itself* at the start of each TBTT [5]. Besides, from Section 2.1, we know that a station which has never received beacons needs to wait at least ProbeDelay before sending *any* frames. Hence by additionally setting ProbeDelay $= \max_{AC \in \{\text{voice, video}\}}\{\text{TXOP}[AC]\} + 2 \times \text{SIFS}$, M-HCCA can lock out the DCF-based access during the CFP even if the beacon frame is sometimes lost.

4) *Loss of control frames.* It is clear that channel errors will degrade the performances of *all* wireless MAC protocols, including M-HCCA. From Appendix D, we can know that, in M-HCCA, misinterpreting a SINGLE handshake result as a COLLISION one due to channel errors in the reservation period may result in, at most, two extra handshakes, the penalty of which is $2 \times (L_{\text{RE}} + L_{\text{RR}})/R_{\min} + \text{PIFS} + 2 \times \text{SIFS}$. Note that even in an error-prone WLAN, the length of the reservation period can be still well controlled by the runtime admission control algorithm. On the other hand, in a round of the polling period, if CF-Poll frames are lost in some sectors but at least one station replies with data, the AP acts as nothing happens since a small amount ($1\% \sim 3\%$) of data loss will not severely degrade the quality of multimedia applications [3]. However, if CF-Poll frames are lost in all sectors and no stations respond, the AP has to poll the set of to-be-polled stations in the next round after an elapsed *PIFS* to prevent the DCF-based interference. Note that when a polled station does not respond to the CF-Poll, the AP infers that the miss-hit problem may occur (even if the fact is not), and the countermeasures are presented in the next Section. Finally, the PL frame is only related to power saving functions and the CF-End is used only to reset the NAV of stations. Hence the loss of PL/CF-End frames only influences the performances and does not affect the correctness of M-HCCA. Fortunately, our experiments reveal that with the aid of PHY rate adaptation, the ratio of data/control frame loss due to channel errors is no more than 1%.

3.2 Beam-Location Related Corner Cases

Since the beam-forming may not be perfect and a station may move during the stream lifetime, the following lists all possible beam-location related problems and their respective detection/resolution methods in M-HCCA.

1) *Beam-overlapping problem, back/side-lobe problem, and unbalanced sector-load problem.* Assume that station A_i is located in the overlapping area of two beams b_k and b_{k+1}. Besides, we assume that station A_j is located very close to the AP; in this case, all beam-sectors can hear the transmission from A_j. In M-HCCA, the AP maintains the beam-location information for each admitted station. During the reservation period, if the AP correctly receives the PR/RR frames from A_i and A_j, the beam-locations of A_i and A_j will be recorded as $\{b_k, b_{k+1}\}$ and $\{b_0, b_1, \ldots, b_{N-1}\}$, respectively. Thus both the beam-overlapping problem and the back/side-lobe problem will not affect the correctness of M-HCCA scheduling algorithm presented in Appendix C since it is *beam-location-aware*. This also implies that the negative effects of unbalanced sector-load problem can be minimized. On the other hand, the beam-overlapping problem and the back/side-lobe problem also do not affect the correctness of the reservation procedure and the polling procedure of M-HCCA. This is because M-HCCA admission control algorithm has taken into account the worst case scenario (e.g. all stations are very close to the AP) under which the reservation procedure and the polling procedure of M-HCCA are in fact, respectively, reduced to that of UPCF [3] and that of HCCA [6]. References [3], [6] have verified the correctness of UPCF and HCCA.

2) *Multipath rich problem.* In multipath rich WLANs, any station may hear transmissions from all sectors; vice versa, all sectors at the AP can hear transmissions from that station. In this case, the beam-location of *every* admitted station will be recorded as $\{b_0, b_1, \ldots, b_{N-1}\}$ and hence it is better for the AP to run in the omni-antenna mode. In M-HCCA, when continuously receiving K (say, $K = 5$) frames from all sectors, the AP then switches to the omni-antenna mode.

3) *Miss-hit problem.* Since the beam-location information cached in the AP may be inaccurate when stations move, the probability of "miss-hit" gets higher with the increase of station mobility. M-HCCA offers a location updating mechanism to minimize the negative effects. Recall that our admission control ensures that each admitted station is polled exactly once during the polling period. Thus once an admitted station A_i does not respond to the CF-Poll, the AP infers that the miss-hit problem may occur. However, the AP does not know its current beam-location. Thus the AP omni-directionally sends the LE (*location enquiry*) frame during the CP to ask station A_i to immediately respond with the LU (*location update*) frame. Once A_i replies with the LU frame, it will automatically refresh the cached beam-location information. However, if A_i does not respond to the LE three times in a row, the AP accordingly removes it from the polling list.

4 Performance Evaluation

4.1 Simulation Models

We follow the event-driven approach [8] to build simulators to compare the performances of M-HCCA to those of existing MAC protocols, i.e., Wang's protocol [19] and Tang's protocol [17]. We assume that, in our considered WLAN, the physical layer is 802.11a, which supports three mandatory PHY rates [4], i.e., 6 Mbps, 12 Mbps, and 24 Mbps. The SNR threshold θ_m for the PHY rate m (Mbps) is shown in Table 1, which summarizes the MAC/PHY parameter values in our simulations. Note that P_{TRANSMIT} denotes the power consumed by the network interface in *transmit* state.

TABLE 1
MAC/PHY Parameters

Parameter	Value
Superframe length	25 ms
SlotTime	9 μs
SIFS	16 μs
$\mathrm{CW}_{\min}$	15 slots
$\mathrm{CW}_{\max}$	1023 slots
RTS/CF-Poll/LE frame size	20 bytes
CTS/ACK/LU frame size	14 bytes
PE frame size	17 bytes
RE frame size	38 bytes
PR/RR frame size	34 bytes
P_{TRANSMIT}	1.65 Watt
P_{RECEIVE}	1.4 Watt
P_{LISTEN}	1.15 Watt
P_{DOZE}	0.045 Watt
SNR threshold θ_6	5 dB
SNR threshold θ_{12}	8 dB
SNR threshold θ_{24}	15 dB
γ	2.56
P_{noise}	-95 dBm
d_0	20 m
P_t	16.2 dBm
G_t	0 dBi
G_r	0 dBi
I_m	5 dB
$\overline{PL}(d_0)$	73 dB

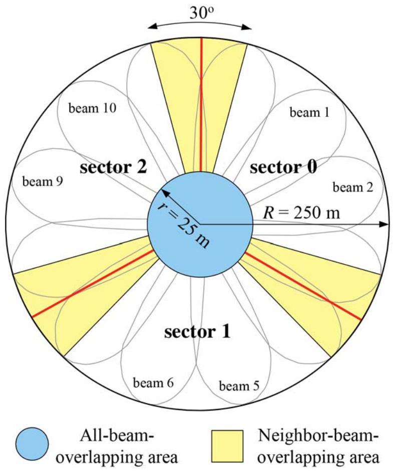


Fig. 9. Simulated imperfect antenna model.

In our simulations, we do not count the energy consumption of the AP since it is often considered to have unlimited power resources.

Our wireless channel model follows the assumptions of [1], [7], [12]. Since the multi-beam AP is suitable for outdoor environments [19], we use the log-distance path loss model [12]. The average path loss for a transmitter-receiver separation d is $\overline{PL}(d) = \overline{PL}(d_0) + 10\gamma \log_{10}(d/d_0)$, where d_0 is the close-in reference distance and γ is the path loss exponent. To estimate $\overline{PL}(d_0)$, we use the Friis free space equation $P_r(d_0) = (P_t G_t G_r \ell^2)/\left[(4\pi)^2 {d_0}^2 L\right]$, where P_t and P_r are the transmit and receive power, and G_t and G_r are the antenna gains of the transmitter and receiver, ℓ is the carrier wavelength, and L is the system loss factor which is set to 1 in our simulations. The received power is $P_r(d) = P_t - \overline{PL}(d)$. Let P_{noise} be the receiver noise power and I_m be the implementation margin. According to [1], the signal-to-noise ratio, $\text{SNR}(d)$, at distance d can be estimated by $\text{SNR}(d) = G_t + G_r + P_r(d) - P_{\text{noise}} - I_m$. Table 1 includes the values of wireless channel parameters in our simulations.

Our antenna model follows the assumptions of [17], [19]. Specifically, we assume that the AP consists of 12 beams with 30° beamwidth per beam; besides, there are $2 \leq M \leq 4$ sectors. By default, we assume that (i) $M = 3$, (ii) the beam-forming is perfect, and (iii) there is no beam-overlapping problem and back/side-lobe problem. However, when studying the imperfect beam-forming scenarios (only in Section 4.7), we consider the antenna model shown in Fig. 9, which abstracts beam-overlapping problem and back/side-lobe problem as neighboring-beam-overlapping problem and all-beam-overlapping problem, respectively. Note that, in our simulations, we do not consider the multipath rich problem since in multipath rich WLANs, M-HCCA is reduced to UPCF and we have evaluated the performances of UPCF in [3].

In our simulations, we focus only on the uplink traffic. Moreover, we consider three kinds of traffic: best-effort data traffic, voice traffic, and video traffic. Each station has only one kind of traffic to send. The data traffic of each best-effort station is modeled by a Poisson process with mean rate λ frames per second. The data frame size is fixed at 1500 bytes. The voice station adopts the scalable audio coder G.729.1 [18] to send the audio stream; thus the voice bit rate can be 8 or $12 + 2k$ Kbps, where $0 \leq k \leq 10$. The video station adopts the scalable video coding scheme to send the three-layer video stream. The bandwidth requirements for sending the base layer, the enhancement layer 1, and the enhancement layer 2 are assumed to be 250, 150, and 100 Kbps, respectively. We set TXOP[voice] = 95 μs and TXOP[video] = 1200 μs. Note that voice and video frames that cannot be transmitted within their respective tolerable delay ($\mathit{delay}_{\text{voice}} = 50$ ms and $\mathit{delay}_{\text{video}} = 75$ ms) will be dropped. For fair comparison, we assume that M-HCCA, Wang's protocol, and Tang's protocol adopt the same rate adaptation schemes.

Two major performance metrics are used in the simulations: the throughput and the *real-time throughput*, which can be also viewed as an indicator of whether a MAC protocol is suitable for multimedia applications. Let $\mathcal{D}$ be the amount of data sent from real-time stations to the AP in delay constraints during the simulation time. In M-HCCA, the real-time throughput [7] is defined as

$$\frac{\mathcal{D}}{\sum_{i=1}^{N_{\text{SF}}} \mathit{time}(\text{CFP}_i)}, \tag{5}$$

where N_{SF} is the total number of superframes during the entire simulation time and $\mathit{time}(\text{CFP}_i)$ is the duration of the i-th contention-free period. On the other hand, since Wang's protocol and Tang's protocol are contention-based, the real-time throughput of their protocols could be defined as

$$\frac{\mathcal{D}}{\sum_{i=1}^{N_{\text{SF}}} \mathit{time}(\text{SF}_i) \times \delta(\text{SF}_i)}, \tag{6}$$

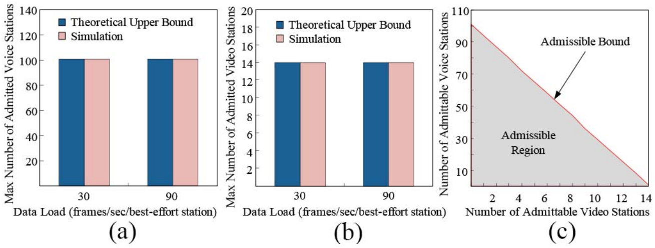


Fig. 10. Number of real-time stations admitted by M-HCCA. (a) All real-time stations are voice stations. (b) All real-time stations are video stations. (c) Admissible region. (The total number of best-effort stations is 30. All stations are static and uniformly distributed in the coverage of the AP.)

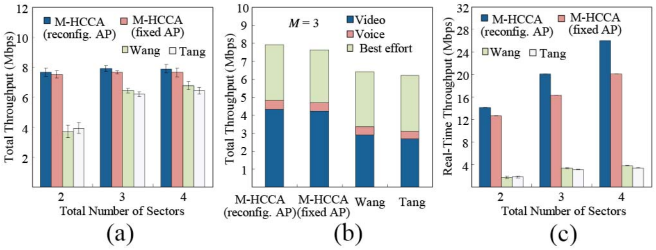


Fig. 11. Performance comparisons under different number of sectors. (a) Total throughput. (b) Anatomy of total throughput. (c) Real-time throughput. ($N_{\text{BE}} = 30$, $N_{\text{voice}} = 22$, and $N_{\text{video}} = 11$. $\lambda = 60$ frames per second. All stations are static and uniformly distributed in the coverage of the AP.)

where $time(\text{SF}_i)$ is the duration of the i-th superframe and

$$\delta(\text{SF}_i) = \begin{cases} 0, & \text{if the AP does not receive real-time} \\ & \text{frames during the } i\text{-th superframe,} \\ 1, & \text{otherwise.} \end{cases}$$

4.2 Admission Control

To verify the accuracy of the run-time admission control algorithm, we measure the capacity of M-HCCA (namely, the maximum number of real-time stations that the AP can admit) under the pure voice/video traffic conditions. According to inequality (4), we can derive that when all real-time stations have the same access category h, the maximum polling list size $\mathcal{L}_h$ is bounded by

$$\mathcal{L}_h \leq \left\lfloor \frac{\text{CFPMaxDuration} - \widehat{T}_s - O_{\text{CFP}}}{\text{TXOP}[h] + T_{\text{CF-Poll}} + 2 \times \text{SIFS}} \right\rfloor. \tag{7}$$

Fig. 10 shows the simulation results. We can see that no matter how the best-effort load varies, the maximum polling list size in M-HCCA exactly matches the theoretical upper bound. These results justify the superiority of our mobile-assisted admission control scheme.

Let N_{voice} and N_{video} denote the numbers of voice stations and video stations, respectively. The *admissible bound* is defined as the combination $(N^*_{\text{voice}}, N^*_{\text{video}})$ of the maximum number of admittable stations in each access category. Let $N^*_{rt} = N^*_{\text{voice}} + N^*_{\text{video}}$ and $\Gamma = \text{CFPMaxDuration} - \widehat{T}_s - O_{\text{CFP}} - N^*_{rt}(T_{\text{CF-Poll}} + 2 \times \text{SIFS})$. Since we require that $\Delta_g \geq 0$, by inequality (4) and the definition of admissible bound, the values of N^*_{voice} and N^*_{video} must satisfy the following two inequalities.

$$N^*_{\text{voice}} \times \text{TXOP[voice]} + N^*_{\text{video}} \times \text{TXOP[video]} \leq \Gamma. \tag{8}$$

$$\begin{aligned} N^*_{\text{voice}} \times \text{TXOP[voice]} + N^*_{\text{video}} \times \text{TXOP[video]} \\ + \min\{\text{TXOP[voice]}, \text{TXOP[video]}\} > \Gamma. \end{aligned} \tag{9}$$

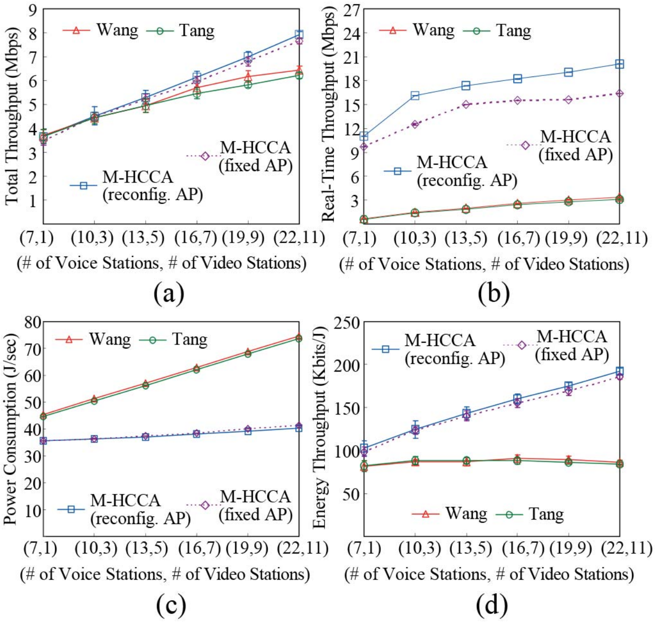
The *admissible region* is defined as the set of ordered pairs $\{(N_{\text{voice}}, N_{\text{video}}) \mid N_{\text{voice}} \leq N^*_{\text{voice}} \text{ and } N_{\text{video}} \leq N^*_{\text{video}}\}$. Fig. 10(c) shows the admissible region under M-HCCA.

4.3 Effect of the Number of Sectors

Fig. 11(a) shows that the throughput of all protocols increases with the increasing number of sectors. This is because when the number of sectors increases, the number of stations that can concurrently send their respective data frames to the AP may increase. However, in M-HCCA, the AP still operates in the omni-antenna mode during the CP. Thus the throughput of M-HCCA rises very slowly when more sectors are employed. Fortunately, for all $2 \leq M \leq 4$, the throughput of M-HCCA is higher than that of two other protocols. Fig. 11(b) shows evidence that the throughput contributed by real-time traffic in M-HCCA is much higher than that in Wang's protocol and Tang's protocol. This is because M-HCCA can reserve the access floor for each admitted real-time station in every superframe. Thus Fig. 11(c) shows that the real-time throughput of M-HCCA can steeply rise as the total number of sectors increases. Clearly, in contrast to the fixed multi-beam AP, the reconfigurable multi-beam AP can help M-HCCA to achieve higher real-time throughput by adaptively adjusting the sector configuration to shorten the collision resolution period and to boost transmission parallelism in the polling period. However, Fig. 11 shows that the throughput differences between M-HCCA with fixed and reconfigurable multi-beam AP are small. This is because in the experiments of Fig. 11, the length of CP is longer than that of CP. From Fig. 11, we also notice that when $M \geq 3$, the throughput of Tang's protocol is slightly lower than that of Wang's protocol. From Appendix A, we know that in Tang's protocol, the AP should sequentially send RTR frames and CTS frames. These rules result in a longer superframe length, thus making the AP drop more real-time frames due to delay expiry.

4.4 Effect of Real-Time Traffic Load

To understand the effect of real-time traffic load, we fix N_{BE} as 30 and vary $(N_{\text{voice}}, N_{\text{video}})$ from $(7, 1)$ to $(22, 11)$. Since Wang's protocol and Tang's protocol do not perform admission control, we only consider the cases where the values of the pair $(N_{\text{voice}}, N_{\text{video}})$ are in the admissible region. Under such conditions, the throughput and the real-time throughput of all protocols can increase with the increasing of real-time traffic load, as shown in Fig 12(a) and (b). However, from Fig. 12(a), we find that when $N_{\text{voice}} \geq 13$ and $N_{\text{video}} \geq 5$, the throughput of Wang's protocol and Tang's protocol increases slower than that of M-HCCA as real-time traffic load increases. The reasons are as follows. Since both Wang's protocol and Tang's protocol are contention-based schemes, when the real-time traffic load becomes heavier, stations are getting harder to contend for the access right, and hence the number of dropped real-time frames due to the violation of delay constraints increases. From Fig. 12(b), we see that M-HCCA has much higher real-time throughput than the other two protocols. As we



(a) (b) (c) (d)

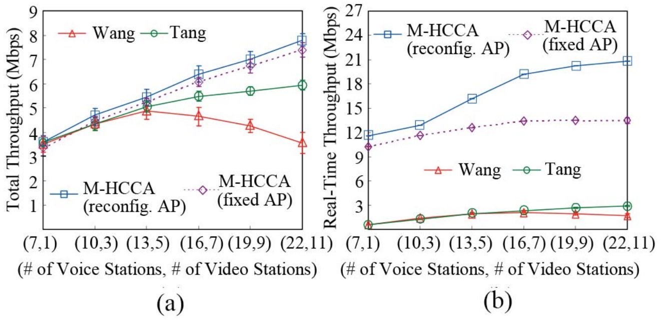
Fig. 12. Performance comparisons under different real-time traffic loads. (a) Total throughput. (b) Real-time throughput. (c) Power consumption. (d) Energy-throughput. ($M = 3$. $N_{\mathrm{BE}} = 30$ and $\lambda = 60$ frames per second. All stations are static and uniformly distributed in the coverage of the AP.)

explained in Section 4.3, this is mainly because M-HCCA is a reservation-based scheme.

Fig. 12(c) shows that the power consumption of Wang's protocol and Tang's protocol is much higher than that of M-HCCA. This is because Wang's protocol and Tang's protocol do not provide any power saving mechanisms, while M-HCCA employs the energy-conserving scheduling such that admitted stations can spend as little awake time as possible. Next, we examine the *energy throughput* [3], which is defined by dividing the amount of data sent from sources to destinations in delay constraints by the total energy consumption of all stations. Evidently, using energy throughput to judge the goodness of a MAC protocol is fairer than using total power consumption since some MAC protocols may consume very little energy, but also achieve very little throughput. Fig. 12(d) shows that M-HCCA has the highest energy throughput.

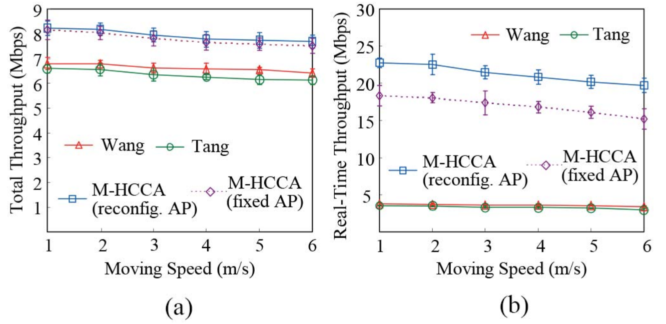
4.5 Effect of Station Distribution

To examine the effect of station distribution, we intentionally assume that all stations are located in the east part of the coverage of the AP. Hence even though we assume that the AP consists of 3 sectors, under this scenario, the number of well-functioning sectors at the fixed multi-beam AP is reduced to 2. Fig. 13(a) and (b) depict that the throughput and the real-time throughput of M-HCCA and Tang's protocol increase monotonically as the real-time traffic load increases. In contrast, the throughput and the real-time throughput of Wang's protocol initially increase as both N_{voice} and N_{video} increase, and then begin to drop when $N_{\mathrm{voice}} > 13$ and $N_{\mathrm{video}} > 5$. This is because Wang's protocol is a contention-based MAC scheme and does not offer any mechanisms to adjust the sector configuration according to station distribution.



(a) (b)

Fig. 13. Performance comparisons in uneven station-distribution environments. (a) Total throughput. (b) Real-time throughput. ($M = 3$. $N_{\mathrm{BE}} = 30$ and $\lambda = 60$ frames per second. All stations are static.)



(a) (b)

Fig. 14. Performance comparisons under varying moving speed. (a) Total throughput. (b) Real-time throughput. ($M = 3$. $N_{\mathrm{BE}} = 30$, $N_{\mathrm{voice}} = 22$, and $N_{\mathrm{video}} = 11$. $\lambda = 60$ frames per second.)

4.6 Effect of Mobility

This Section evaluates the performance of M-HCCA, Wang's protocol and Tang's protocol under various degrees of station mobility. In the experiments, we consider the random way-point model, in which all stations alternate between pausing and then move to a randomly chosen location (in the coverage of the AP) at a fixed speed. The pause time is fixed at 30 s. Fig 14(a) and (b) show that the throughput and the real-time throughput of these three protocols monotonically decrease as the moving speed of stations increases. This is because mobility may sometimes result in a situation where stations are unevenly distributed among all sectors. However, we notice that, with the increase of station mobility, the real-time throughput of M-HCCA degrades more significantly than that of the other two protocols. The reasons are as follows. In Wang's protocol and Tang's protocol, the miss-hit events can occur only in the downlink access. However, M-HCCA is a polling-based scheme. Thus the miss-hit events may also happen in the uplink access and occur more often in a higher mobility environment.

4.7 Effect of Imperfect Beam-Forming

Imperfect beam-forming does lead to a performance loss by reducing independent spatial reuse area, as shown in Fig. 15, where "Wang+imperfect" denotes that Wang's protocol runs in imperfect beam-forming environments. Fig. 15(a) and (b) show that the throughput difference between "Wang+perfect" and "Wang+imperfect" is much larger than that between "Tang+perfect" and "Tang+imperfect." This is because in Wang's protocol,

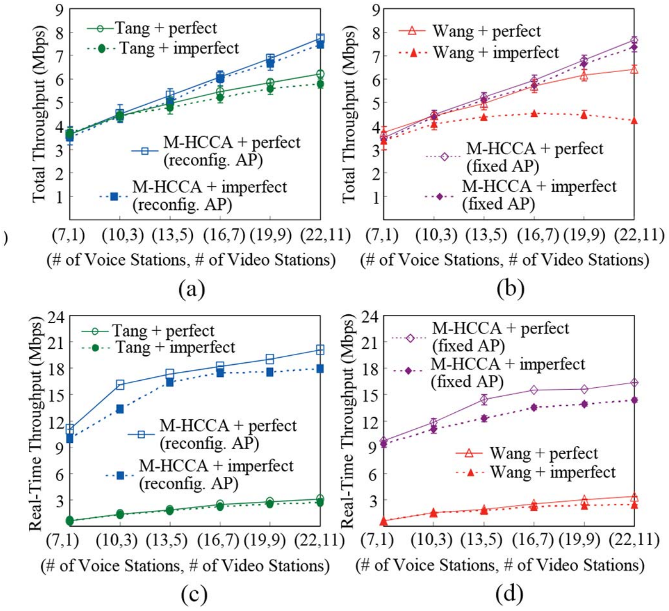


Fig. 15. Performance comparisons under perfect/imperfect beam-forming environments. (a) Tang's protocol v.s. M-HCCA with regard to total throughput. (b) Wang's protocol v.s. M-HCCA with regard to total throughput. (c) Tang's protocol v.s. M-HCCA with regard to real-time throughput. (d) Wang's protocol v.s. M-HCCA with regard to real-time throughput. (All stations are static and uniformly distributed in the coverage of the AP. $M = 3$. $N_{\rm BE} = 30$ and $\lambda = 60$ frames per second.)

once the AP detects the beam-overlapping problem and back/side-lobe problem, it resolves these problems by sequentially replying with CTS frames sector-by-sector to *each* station that has successfully sent the RTS. This method significantly lengthens the superframe length, thus increasing the number of dropped real-time frames. Fig. 15(c) and (d) show that there is a real-time throughput gap between "M-HCCA+perfect" and "M-HCCA+imperfect" since in M-HCCA, imperfect beam-forming leads to a longer reservation period and a reduction of transmission parallelism in the polling period. However, Fig. 15(a) and (b) show that the throughput difference between "M-HCCA+perfect" and "M-HCCA+imperfect" is small since in M-HCCA, the AP operates in the omni-antenna mode during the CP, and the length of CP is longer than that of CFP in the experiments.

5 Conclusion

Theoretically, the capacity of a WLAN can be considerably boosted by the use of multi-beam smart antennas. However, if we directly apply 802.11 to a WLAN with multi-beam AP, we will inevitably encounter many challenges, including receiver blocking problem, unnecessary defer problem, beam-overlapping problem, back/side-lobe problem, hidden terminal problem, multipath rich problem, and miss-hit problem. The existing solutions [17], [19] to these problems are based on the DCF and hence not suitable for multimedia applications. In this paper, we have proposed a novel polling-based MAC protocol, named M-HCCA, for a WLAN with multi-beam AP. What makes M-HCCA so versatile and unique is that it not only resolves all the above-mentioned problems in a *simple and effective* manner, but also *integrates* non-reversal prioritization, time-bounded reservation, cross-layer rate adaptation, energy-conserving scheduling, and mobile-assisted admission control into *one* scheme to support real-time multimedia traffic. Extensive simulation results do confirm that, in terms of throughput, real-time throughput, and energy throughput, M-HCCA significantly outperforms existing protocols [17], [19] even in uneven station distribution, imperfect beam-forming, and high mobility environments.

Acknowledgments

The authors would like to thank the editor and the reviewers for their valuable comments that helped to improve the quality of this paper. This work was supported in part by the National Science Council, Taiwan, under grant NSC 98-2221-E-110-048.

References

[1] O. Awoniyi and F. A. Tobagi, "Effect of fading on the performance of VoIP in IEEE 802.11a WLANs," in *Proc. IEEE Int. Conf. Communications*, vol. 6. pp. 3712–3717, Jun. 2004.

[2] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.

[3] Z.-T. Chou, C.-C. Hsu, and S.-N. Hsu, "UPCF: A new point coordination function with QoS and power management for multimedia over wireless LANs," *IEEE/ACM Trans. Netw.*, vol. 14, no. 4, pp. 807–820, Aug. 2006.

[4] M. S. Gast, *802.11 Wireless Networks: The Definitive Guide*, 2nd ed. Beijing, China: O'Reilly Inc., 2005.

[5] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, Nov. 1999.

[6] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*. IEEE Standard 802.11e, Nov. 2005.

[7] B. Kim, S. Kim, Y. Fang, and T. F. Wong, "Two-step multipolling MAC protocol for wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 6, pp. 1276–1286, Jun. 2005.

[8] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed. New York, NY, USA: McGrraw-Hill, 2000.

[9] Noterl. (2008). *Wireless Mesh Network:Outdoor Wi-Fi Made Simple* [Online]. Available: http://pdf.aminer.org/000/366/431/dimensioning_of_wireless_links_sharing_voice_and_data.pdf

[10] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 3rd ed. New York, NY, USA: Springer, 2008.

[11] *Plasma Antennas*[Online] Available: http://www.plasmaantennas.com

[12] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[13] H. Schwarz , D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[14] I. Stevanović, A. Skrivervik, and J. R. Mosig. (2003). Smart antenna systems for mobile communications. Ecole Polytechnique Federale De Lausanne, Lausanne, Switzerland [Online]. Available: http://infoscience.epfl.ch/record/140902/files/smart_antennas.pdf.

[15] J. A. Stine and G. D. Veciana, "Improving energy efficiency of centrally controlled wireless data networks," *ACM/Kluwer Wireless Netw.*, vol. 8, pp. 681–700, Nov. 2002.

[16] K. Sundaresan and R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 546–559, Jun. 2007.

[17] Z. Tang, X. Xing, and F. Jiang, "Providing balanced and enhanced transmission for WLANs with multi-beam access point," in *Proc. IEEE Communication Networks Services Research Conf.*, Halifax, NS, Canada, 2008, pp. 242–248.

[18] I. Varga, S. Proust, and H. Taddei, "ITU-T G.729.1 scalable codec for new wideband services," *IEEE Commun. Mag.*, vol. 47, no. 10, pp. 131–137, Oct. 2009.

[19] J. Wang, Y. Fang, and D. Wu, "Enhancing the performance of medium access control for WLANs with multi-beam access point," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 556–565, Feb. 2007.

[20] D.-Y. Yang, T.-J. Lee, K. Jang, J.-B. Chang, and S. Choi, "Performance enhancement of multirate IEEE 802.11 WLANs with geographically scattered stations," *IEEE Trans. Mobile Comput.*, vol. 5, no. 7, pp. 906–919, Jul. 2006.

Zi-Tsan Chou received the Ph.D. degree in Computer Science and Information Engineering from National Taiwan University. He is currently an associate professor in the Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan. His industrial experience includes NEC Research Institute (America) and Institute for Information Industry (Taiwan). He holds eight United State patents, eleven Taiwan patents, and three China patents in the field of wireless communications. His research interests include medium access control, power management, and quality-of-service control for wireless networks. He is a member of both the IEEE and the IEEE Communications Society.

Cong-Qi Huang received the BS degree in Computer Science and Information Engineering from National United University, Miaoli, Taiwan, and the MS degree in Electrical Engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan in 2008 and 2011, respectively. He is currently a firmware engineer at the ACARD Technology Corp., Kaohsiung. His current research interests include wireless medium access control and multimedia applications.

J. Morris Chang received the PhD degree from North Carolina State University. He is an associate professor at Iowa State University. His industrial experience includes Texas Instruments and AT&T Bell Labs. His current research interests include computer security, wireless networks, and computer systems. Currently, he is a handling editor of the *Journal of Microprocessors and Microsystems* and the Associate Editor-in-Chief of *IEEE IT Professional*. He is a senior member of the IEEE.