Wireless Netw (2009) 15:443–454
DOI 10.1007/s11276-007-0062-5

Supporting voice and video applications over IEEE 802.11n WLANs

Lin X. Cai · Xinhua Ling · Xuemin (Sherman) Shen · Jon W. Mark · Lin Cai

Published online: 5 September 2007

Abstract In this paper, an analytical model is developed for the performance study of an IEEE 802.11n wireless local area network (WLAN) supporting voice and video services, considering the new features of the medium access control (MAC) protocol proposed in IEEE 802.11n, i.e., frame aggregation and bidirectional transmission. We show that these enhanced MAC mechanisms can effectively improve the network capacity by not only reducing the protocol overheads, but also smoothing the AP-bottleneck effect in an infrastructure-based WLAN. Voice and video capacity under various MAC mechanisms are compared as well.

Keywords IEEE 802.11n · Frame aggregation · Bidirectional transmission · Voice capacity · Video capacity

1 Introduction

IEEE 802.11 wireless local area networks (WLANs) have been widely deployed for wireless Internet access. The legacy 802.11, 802.11b, 802.11a/g can provide up to 2 Mbps, 11 Mbps, and 54 Mbps data rates, respectively. However, the achievable throughput of a WLAN is less than half of the physical layer (PHY) raw data rate because of the protocol overheads, including protocol headers (e.g., UDP, TCP, IP, medium access control (MAC)), PHY preamble, various inter-frame spaces (IFSs), acknowledgment (ACK) and backoff time, etc. In addition, the majority of existing WLANs are set up in infrastructure mode, where mobile stations access the Internet through an access point (AP). As shown in Fig. 1, the AP delivers all traffic to and from the WLAN, and thus has a much higher traffic load and is the bottleneck.

Although current WLAN applications are mainly data centric, there is a growing demand for bandwidth-intense and delay-sensitive multimedia services over WLANs. Voice capacity, in terms of maximum number of voice connections that can be supported in an IEEE 802.11b WLAN with satisfactory user-perceived quality, has been actively investigated in the literature [1–3]. Due to the inherent protocol inefficiency and AP-bottleneck effect, only a limited number of voice calls can be supported with the DCF-based MAC. In response to the demand for higher performance WLANs to support multimedia applications such as voice, video telephony, video conferencing and high-definition television (HDTV), the 802.11n task group has been established to standardize the next generation WLAN to provide over 100 Mbps throughput at the MAC data service access point (SAP) via PHY and MAC enhancements. An IEEE 802.11n WLAN can operate with physical layer raw data rate up to 200–600 Mbps by using multiple-input multiple-output (MIMO) technology, modified encoding and optional channel binding schemes. To efficiently improve the SAP throughput, two main MAC enhancement mechanisms have been proposed to reduce

L. X. Cai (✉) · X. Ling · Xuemin (Sherman) Shen · J. W. Mark
Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1
e-mail: lcai@bbcr.uwaterloo.ca

X. Ling
e-mail: x2ling@bbcr.uwaterloo.ca

Xuemin (Sherman) Shen
e-mail: xshen@bbcr.uwaterloo.ca

J. W. Mark
e-mail: jwmark@bbcr.uwaterloo.ca

L. Cai
Department of Electrical & Computer Engineering, University of Victoria, Victoria, BC, Canada V8W 3P6
e-mail: cai@ece.uvic.ca

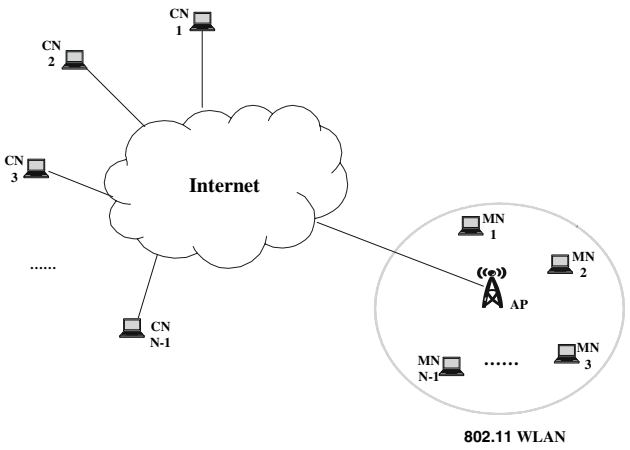


Fig. 1 Network scenario

the protocol overheads: (1) frame aggregation and (2) bidirectional transmission [4]. These mechanisms eliminate the need to initiate a transmission for every MAC frame (MF) in the legacy 802.11, and thus reduce the transmission overheads and improve the throughput efficiency.

The performance of legacy IEEE 802.11 MAC has been extensively studied in the literature [3–6]. Some new mechanisms, i.e., concatenation and piggyback, are proposed and analyzed in [7] under the best-case scenario and the saturation scenario. In [8], saturation throughput of an IEEE 802.11 WLAN with the support of bidirectional frame aggregation is analyzed. Studies in [7, 8] show that these two schemes can greatly improve the system performance under the saturation scenario. However, stations carrying practical applications, e.g., real-time voice and video services, are unsaturated. To the best of our knowledge, there is little work investigating the network capacity, in terms of the number of non-persistent traffic flows that can be supported in an IEEE 802.11n WLAN.

In this paper, we first develop an analytical model to study the network capacity of an IEEE 802.11 WLAN supporting asymmetric, non-persistent traffic, using the carrier sense multiple access with collision avoidance (CSMA/CA) mechanism. We show that the AP becomes the bottleneck under high traffic load since all stations have the same priority to access the channel with the legacy 802.11. Using the enhanced MAC mechanisms in IEEE 802.11n, i.e., downlink frame aggregation and bidirectional transmission, the AP-bottleneck effect can be significantly mitigated. The heavily loaded AP does not need to initiate a transmission for every MF, which improves the transmission efficiency by reducing the number of channel access contentions and the associated overheads. Two new aggregation schemes are also proposed and compared with the existing schemes. We then extend the proposed model to investigate the performance of these enhanced MAC mechanisms. Although this paper studies only a few frame aggregation schemes, the approach is readily extensible to other physical level and MAC level aggregation schemes. To substantiate the analysis, we calculate voice and video capacity of an 802.11n WLAN under various enhanced MAC mechanisms.

The remainder of the paper is organized as follows. The IEEE 802.11n MAC enhancements are introduced in Sect. 2. The system model is presented in Sect. 3. An analytical model for studying the AP-bottleneck effect is developed in Sect. 4.1. The model is then extended for performance analysis of two MAC mechanisms: frame aggregation and bidirectional transmission in Sect. 4.2 and Sect. 4.3, respectively. Section 5 presents the voice and video capacity obtained from the developed model, which is verified through simulations, followed by concluding remarks in Sect. 6.

2 MAC layer enhancements in IEEE 802.11n

In IEEE 802.11n, several new MAC features have been proposed to improve throughput efficiency. One approach is frame aggregation, the idea of which is to aggregate multiple MAC/PHY frames into a single frame (or a train of frames) for transmission [9]. Generally, aggregation mechanisms can be classified from many different aspects: uplink versus downlink, PHY-level versus MAC-level, immediate ACK versus delayed ACK, single-destination versus multi-destination, etc.

Some frame aggregation mechanisms are illustrated in Fig. 2. In Fig. 2(a), a train of N PHY frames are sent one by one with no IFS. These frames can be transmitted to one or multiple destinations, and each destination station acknowledges the received frame in the same order after a short IFS (SIFS). In Fig. 2(b), each destination station sends an ACK immediately after a SIFS when it successfully receives a frame.

Maximizing throughput may require a large aggregation frame with length longer than that specified in the current standard (4095 bytes) [9]. On the other hand, it is suggested that the total length of the aggregation frame be smaller than a threshold since some huge frames may cause unfairness among stations. In addition, long data frames will result in large collision time and thus reduce the transmission efficiency when collision probability is high. In legacy 802.11, the optional request to send/clear to send (RTS/CTS) is proposed to improve the transmission efficiency when the frame size is larger than a threshold (0–2347 bytes). However, RTS/CTS in legacy 802.11 is employed by a pair of sender and receiver for unicast transmission and is not suitable for the downlink aggregation mechanism which may involve multiple destination stations. Therefore, we propose

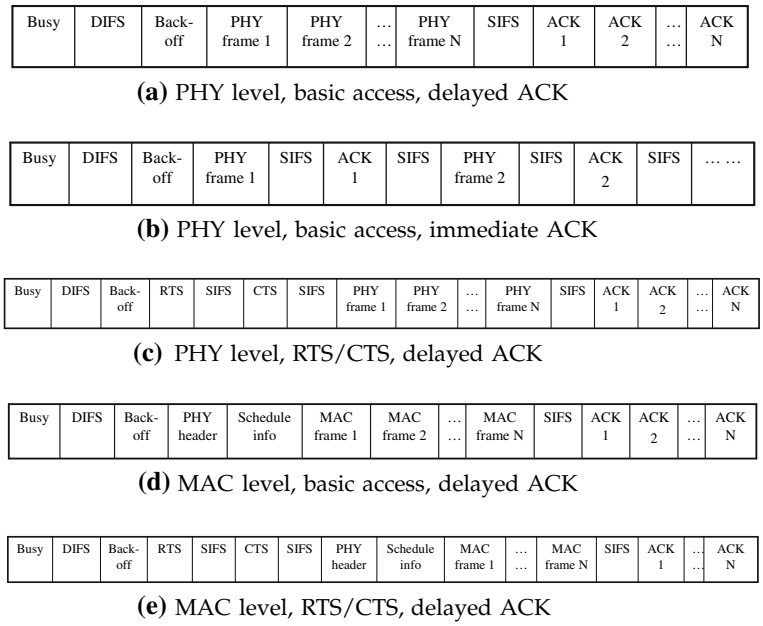


Fig. 2 Frame aggregation mechanisms

a modified RTS/CTS function that can be used with downlink aggregation to reduce collisions resulting from large data frames, as shown in Fig. 2(c). Generally, an RTS frame can be sent in a multicast fashion and all involved destination stations need to send back CTS frames if they are available to receive data. In a single-hop WLAN with no hidden terminals, a modified RTS frame with aggregation information, e.g., a list of destinations and transmission sequence, can be sent out and the destination stations send back CTS frames in the same sequence of the destination list. Upon receiving the RTS frame, all stations check the destination list. Stations not on the list set the network allocation vector (NAV) and will not access the channel during the period indicated by NAV. To further reduce the CTS overhead, another option is that only a couple of destination stations, which are chosen randomly by the transmitter, send back CTS frames in a predetermined sequence.

The above three mechanisms are PHY level aggregations. The PHY overhead can be further reduced through MAC level aggregations, which are shown in Fig. 2(d) and 2(e) for basic access mode and RTS/CTS mode, respectively. With these two mechanisms, N MAC frames for different destinations can be aggregated into one PHY frame. After the (shared) PHY preamble and header, destination stations receive the scheduling information, based on which they can determine the time to receive the MFs if there is any. Using downlink multi-destination aggregation, the AP only needs to contend once to transmit an aggregated frame to multiple MNs, in contrast to multiple contentions and transmissions without frame aggregation.

Another approach to improve throughput efficiency is by allowing data transmission in both directions. That is, a receiver can piggyback aggregation frames to the transmitter without initiating a new transmission, as shown in Fig. 3(a). In an infrastructure WLAN, the AP can piggyback a frame to an MN in the downlink after the MN successfully transmits a frame to the AP in the uplink, and vice versa. Thus, the number of contentions in the WLAN can be significantly reduced and network throughput and capacity will be effectively improved. The bidirectional transmission mechanism is efficient when the traffic flows between the transmitter and receiver are symmetric, but it may not be useful for some other applications, e.g., half-duplex voice services with silence suppression. In other words, bidirectional transmission cannot improve the transmission efficiency if there is no frame in the reverse direction for piggyback. Thus, we extend

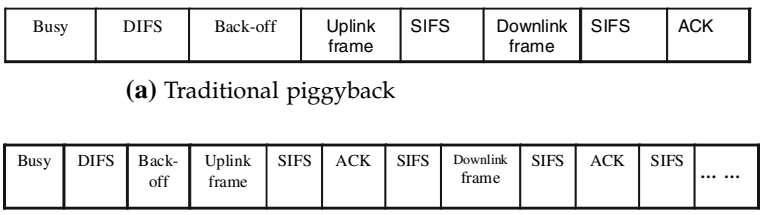


Fig. 3 Bidirectional transmission

the bidirectional transmission scheme in a more general scenario to smooth the AP-bottleneck effect. As shown in Fig. 3(b), the heavily loaded AP can transmit one frame to any destination station without contentions upon receiving an uplink frame.

3 System model

We consider a single-hop fully-connected WLAN with N stations, including one AP and $N-1$ mobile nodes (MNs) as shown in Fig. 1. The AP and $N-1$ correspondent nodes (CNs) are connected via a backbone network. Data or multimedia connections are established between MNs and CNs, through the AP. Since current backbone network is rapidly upgraded to terabit speed, the last-mile wireless access network is usually considered as the bottleneck for achieving the full potential of high speed Internet access. Therefore, in this paper, we focus on the performance analysis of uplink and downlink transmissions in the WLAN. We investigate the capacity of a WLAN supporting wireless multimedia services, e.g., voice, video telephony, video conferencing. We assume all stations in the WLAN can sense the status of the shared wireless channel. Time is discretized into slots, and all stations are synchronized to operate in slotted time. The wireless channel is assumed ideal such that all transmitted frames can be received error-free if there is no collision.

4 Analytical model

4.1 Analytical model with asymmetric traffic

Denote the traffic arrival rate and frame service rate of station i as λ_i and μ_i frames per slot, respectively. The traffic intensity (or queue utilization ratio) of station i is $\rho_i = \lambda_i/\mu_i$. Define p_i the conditional collision probability of frames transmitted by station i and τ_i the transmission probability of station i at the beginning of a randomly chosen time slot. Given station i transmits in a given slot, a collision occurs if at least one of the remaining stations also transmits. We have

$$p_i = 1 - \prod_{j=0, j\neq i}^{N-1} (1 - \rho_j \tau_j), \tag{1}$$

where $i = 0, ..., N-1$. The exponential backoff procedure of CSMA/CA can be modeled as a truncated geometrical random variable, and the average backoff time of station i is derived as [3]

$$E[W_i] = \sum_{k=0}^{m-1} p_i^k (1-p_i) \sum_{j=0}^{k} \frac{CW_j}{2} + p_i^m \sum_{j=0}^{m} \frac{CW_j}{2}, \tag{2}$$

where CW_j is the contention window in the j-th backoff stage and m is the retry limit. During the period of $E[W_i]$, station i makes A_i transmission attempts, which can also be modeled as a truncated geometrical random variable with mean

$$E[A_i] = \sum_{k=0}^{m-1} p_i^k (1-p_i)(k+1) + p_i^m (m+1) = \frac{1-p_i^{m+1}}{1-p_i}. \tag{3}$$

Therefore, the transmission probability τ_i is derived as

$$\tau_i = \frac{E[A_i]}{E[W_i] + E[A_i]}. \tag{4}$$

Substituting (2) and (3) into (4), τ_i is obtained as a function of p_i. Denote T_S and T_C as the time durations for a successful transmission and a collision, respectively. In the basic access mode, $T_S = (P_{OH} + T_{MF}) + SIFS + T_{ACK} + DIFS$ and $T_C = (P_{OH} + T_{MF}) + ACK_{timeout} + DIFS \approx T_S$, where P_{OH} is the PHY overheads including the preamble and physical layer header and T_{MF} is the transmission time of the MAC payload. The number of collisions of a frame can be modeled as a geometrically distributed variable, and the average collision time a frame of station i experienced is $p_i T_C/(1-p_i)$ [3]. During the time interval $1/\mu_i$, from the time station i attempts to transmit a frame until the frame is transmitted successfully, one of the following events must occur: (1) successful transmissions made by the remaining stations; (2) collisions[1]; (3) channel idleness when station i is in the backoff stages. Therefore, we have

$$\frac{1}{\mu_i} = T_{s_i} + \frac{1}{\mu_i} \sum_{j=0, j\neq i}^{N-1} \lambda_j T_{s_j} + E[W_i] + \frac{1}{2}\left(\frac{1}{\mu_i} \sum_{j=0, j\neq i}^{N-1} \lambda_j \overline{T_{c_j}} + \overline{T_{c_i}}\right), \tag{5}$$

where $i = 0,1,\ldots,N-1$.

Given the arrival rates $\vec{\lambda} = [\lambda_0, \lambda_1, \ldots, \lambda_{N-1}]$, the equation sets (1) and (5) can be solved numerically to obtain $\vec{p} = [p_0, p_1, \ldots, p_{N-1}]$, $\vec{\mu} = [\mu_0, \mu_1, \ldots, \mu_{N-1}]$, and $\vec{\rho} = [\rho_0, \rho_1, \ldots, \rho_{N-1}]$. The maximum number of traffic flows that can be supported is obtained when $\rho_i < 1$ for any station i. If $\rho_i \geq 1$, the queue of frames at station i

[1] Because the probability of three or more stations simultaneously transmitting is very small, we assume that collisions are due to two stations transmitting simultaneously.

will build up and all flows of station i will suffer from the ever increasing queuing delay and buffer overflow.

4.2 Analytical model with downlink frame aggregation

Assuming the traffic loads of the MNs are homogeneous and the flows between the AP and MNs are symmetric, the traffic arrival rate of the AP ($i = 0$) is $N-1$ times that of an MN ($i = 1,\ldots,N-1$), i.e., $\lambda_0 = (N-1)\lambda_1$ and $\lambda_i = \lambda_1$, for $\forall i$, $i = 1,\ldots,N-1$. Equations (1) and (5) can be rewritten as

$$p_0 = 1-(1-\rho_1\tau_1)^{N-1} \tag{6}$$

$$p_1 = 1-(1-\rho_1\tau_1)^{N-2}(1-\rho_0\tau_0) \tag{7}$$

$$\frac{1}{\mu_0} = \left((N-1)\frac{\lambda_1}{\mu_0}+1\right)T_S + E[W_0] + \frac{1}{2}\left[(N-1)\frac{\lambda_1}{\mu_0}\frac{p_1 Tc}{1-p_1} + \frac{p_0 T_C}{1-p_0}\right] \tag{8}$$

$$\frac{1}{\mu_1} = \left((N-2)\frac{\lambda_1}{\mu_1}+1+\frac{\lambda_0}{\mu_1}\right)T_S + E[W_1] + \frac{1}{2}\left(\left((N-2)\frac{\lambda_1}{\mu_1}+1\right)\frac{p_1 T_C}{1-p_1} + \frac{\lambda_0}{\mu_1}\frac{p_0 T_C}{1-p_0}\right), \tag{9}$$

where $E[W_0]$ and $E[W_1]$, and τ_0 and τ_1 can be derived from (2) and (4), respectively.

With the legacy DCF MAC, all stations have the same priority for channel access. This is unfavorable to the heavily loaded AP, which becomes unstable when $\rho_0 = (N-1)\lambda_1/\mu_0 \geq 1$. The arrival rate of the AP, λ_0, can be effectively reduced using aggregation mechanisms. We consider that the AP aggregates $N-1$ frames for multiple MNs, using the aggregation schemes shown in Fig. 2. Therefore, λ_0 is equivalent to λ_i, $i = 1,\ldots,N$. Denote the frame transmission time in the downlink and uplink as T_D and T_U, respectively. We have $T_U = DIFS + P_{OH} + T_{MF} + T_{ACK} + SIFS$, $T_D = DIFS + (N-1)(P_{OH} + T_{MF} + T_{ACK}) + SIFS$ for aggregation scheme (a), and $T_D = DIFS + (N-1)(P_{OH} + T_{MF} + T_{ACK}) + 2(N-1)SIFS$ for aggregation scheme (b). During $1/\mu_1$, the AP spends T_D seconds transmitting an aggregation frame and MNs transmit $(N-1)\lambda_1/\mu_0$ frames which contribute $(N-1)\lambda_1 T_U/\mu_0$ seconds. Due to the different lengths of downlink and uplink frames, the collision times of downlink frames T_{CD} and uplink frames T_{CU} are different. The average collision time a downlink frame experiences is

$$\overline{T_{CD}} \approx \frac{p_0}{1-p_0}T_D. \tag{10}$$

An uplink frame may collide with another uplink frame with probability $(N-2)\rho_1/[(N-2)\rho_1+\rho_0]$ and a downlink frame with probability $\rho_0/[(N-2)\rho_1+\rho_0]$. The average collision time an uplink frame experiences is

$$\overline{T_{CU}} \approx \frac{p_1}{1-p_1}\left(\frac{(N-2)\rho_1}{(N-2)\rho_1+\rho_0}T_U + \frac{\rho_0}{(N-2)\rho_1+\rho_0}T_D\right). \tag{11}$$

Equations (8) and (9) can thus be rewritten as

$$\frac{1}{\mu_0} = (N-1)\frac{\lambda_1}{\mu_0}T_U + T_D + \frac{1}{2}\left((N-1)\frac{\lambda_1}{\mu_0}\overline{T_{CU}} + \overline{T_{CD}}\right) + E[W_0], \tag{12}$$

$$\frac{1}{\mu_1} = \left((N-2)\frac{\lambda_1}{\mu_1}+1\right)T_S + \frac{\lambda_1}{\mu_1}T_D + E[W_1] + \frac{1}{2}\left(\left((N-2)\frac{\lambda_1}{\mu_1}+1\right)\overline{T_{CU}} + \frac{\lambda_1}{\mu_1}\overline{T_{CD}}\right). \tag{13}$$

With CSMA/CA, stations have to wait T_C each time a collision occurs. In the basic access mode, long frames results in large T_C, which degrades transmission efficiency significantly when the collision probability is high. The RTS/CTS mode is an option used by the legacy MAC to reduce data frame collisions when the frame payload exceeds a threshold. As shown in Fig. 2(c), in the RTS/CTS mode, $T_D = DIFS + RTS + CTS + (N-1)(P_{OH} + T_{MF} + ACK) + 3SIFS$. A CTS timeout implies a collision and the transmitter will re-initiate a transmission following the CSMA/CA mechanism. Since RTS and CTS frames are very small, the duration of a collision depends on the uplink data frame. When the frame length does not exceed the RTS threshold, the uplink frames are transmitted in the basic access mode, so $T_C \approx T_U$; otherwise, the uplink frames are transmitted in the RTS/CTS mode, and $T_C = RTS + CTS_{timeout} + DIFS$. Then, Eqs. (12) and (13) are simplified as

$$\frac{1}{\mu_0} = (N-1)\frac{\lambda_1}{\mu_0}T_U + T_D + E[W_0] + \frac{1}{2}\left((N-1)\frac{\lambda_1}{\mu_0}\frac{p_1 T_C}{1-p_1} + \frac{p_0 T_C}{1-p_0}\right) \tag{14}$$

$$\frac{1}{\mu_1} = \left((N-2)\frac{\lambda_1}{\mu_1}+1\right)T_U + \frac{\lambda_1}{\mu_1}T_D + E[W_1] + \frac{1}{2}\left(\left((N-2)\frac{\lambda_1}{\mu_1}+1\right)\frac{p_1 T_C}{1-p_1} + \frac{\lambda_1}{\mu_1}\frac{p_0 T_C}{1-p_0}\right). \tag{15}$$

The PHY overheads can be further reduced with a MAC-level aggregation scheme. As shown in Fig. 2(d), $T_D = DIFS + P_{OH} + T_{sch} + (T_{MF} + T_{ACK})(N-1) + SIFS$,

where T_{sch} is the transmission time of the scheduling information that is transmitted before the MFs. After synchronization, stations can determine when to receive their MFs by checking the scheduling information, instead of receiving the aggregated PHY frame. Hence, strict synchronization is critical for the MAC aggregation scheme. Similar to Fig. 2(b), RTS/CTS can also be used in MAC aggregation for transmission efficiency, and T_C can be either $T_C \approx T_U$ or $T_C = RTS + CTS_{timeout} + DIFS$. Substitute T_D and T_C into (14) and (15), we can obtain ρ_i of station i.

4.3 Analytical model with bidirectional transmission

We consider an infrastructure WLAN carrying symmetric traffic flows between the AP and MNs. After a station successfully transmits a frame to the AP, the AP can piggyback data frames to the transmitter without initiating a new transmission, and vice versa. Therefore, half of the frames are piggybacked without contentions, while all frames have to be transmitted via contention using the legacy DCF MAC. We develop an analytical model for the downlink piggyback scheme, i.e., the AP that carries half of the frames in the WLAN do not contend with the MNs. Thus, only the $N-1$ MNs contend for uplink transmissions. We have

$$p_1 = 1 - (1 - \rho_1\tau_1)^{N-2}, \tag{16}$$

$$\frac{1}{\mu_1} = \left((N-2)\frac{\lambda_1}{\mu_1} + 1\right)T_S + \frac{1}{2}\left((N-2)\frac{\lambda_1}{\mu_1} + 1\right) \times \frac{p_1 T_C}{1-p_1} + E[W_1], \tag{17}$$

where τ_1 is a function of p_1 derived from (4). When the traditional piggyback is used, as shown in Fig. 3(a), $T_S = DIFS + 2(P_{OH} + T_{MF}) + 2SIFS + T_{ACK}$ and $T_C = DIFS + 2(P_{OH} + T_{MF}) + SIFS$. In the extended scheme, as shown in Fig. 3(b), $T_S = DIFS + 2(P_{OH} + T_{MF}) + 3SIFS + 2T_{ACK}$ and $T_C = DIFS + P_{OH} + T_{MF} + SIFS$.

5 Capacity evaluation of an IEEE 802.11n WLAN

To substantiate the analysis, we calculate the capacity of an IEEE 802.11n WLAN supporting multimedia services, such as low rate voice and broadband video applications. Generally, in a voice/video over IP (VoIP) system, analogue signals are first digitized, compressed and encoded into digital voice/video streams by the codecs. The output streams are then packetized for efficient and network-friendly transmissions over an IP-based network [10, 11]. In general, multimedia streams are encapsulated with RTP/UDP/IP headers. After the voice/video packets are delivered through the network, the reverse processes of decoding and depacketizing is accomplished at the receiver.

We use Maple 9.5 [12] to calculate the analytical results and validate them through extensive simulations with an event-driven simulator written in the C language. The system parameters are listed in Table 1. The simulations employ the system model as in Sect. 3. To show the queue accumulating effect in the AP, we set the buffer size of the AP to 300 packets. In the initial stage, a voice or a video connection is established during every codec sample period to gradually approach the network capacity, with the starting time randomly chosen over the sample period. To eliminate the warming-up effects, the simulation data are collected from 20 s to 200 s.

5.1 Voice capacity evaluation

The main attributes of some frequently used voice codecs are listed in Table 2. Different codecs use different compression algorithms resulting in different bit rates. G.711 is the international standard for encoding telephone audio, which has a fixed bit rate of 64 Kbps. With a 10 ms sample period, corresponding to a rate of 100 packets per second, the payload size is 64000/(100 * 8) = 80 bytes. When the sample period is increased to 20 ms, corresponding to 50 packets per second, the payload size is increased to 160 bytes accordingly. Compared to G.711, G.723 and G.729 have lower bit rates at a cost of higher codec complexity. G.723 is one of the most efficient codecs with the highest compression ratio, and is usually used in video conferencing applications. G.729 is an industry standard with high bandwidth utilization for toll quality voice calls.

Table 1 System parameters

PHY layer data rate	216 Mbps
PHY preamble & header (P_{OH})	24 μs
DIFS	34 μs
SIFS	16 μs
A slot time	9 μs
CWmin	16
CWmax	512
Retry limit m	7
MAC header & FCS	34 bytes
IP/UDP/RTP header	40 bytes
T_{ACK}, CTS	24.5 μs
RTS	24.7 μs
T_{sch}	300 bytes

Table 2 Frequently used voice codecs

Voice codec		G.711	G.723a	G.729
Codec bit rate (Kbps)		(64)	(5.3/6.3)	(8)
Sample Period (ms)	Arrival rate (frames/s)	Payload (byte)	Payload (byte)	Payload (byte)
10	100	80		10
20	50	160		20
30	33.33	240	20/24	30
40	25	320		40
50	20	400		50

When we calculate the voice capacity of an infrastructure WLAN with the developed models, the voice traffic is considered as a constant bit rate (CBR) flow because of three reasons: (1) many voice codecs do not use silence suppression schemes; (2) even if silence suppression is used and voice traffic exhibits on-off characteristics, some packets containing background noise are transmitted intermittently during "off" periods to obtain a better voice quality [13]; (3) a tighter bound derived with CBR traffic is robust in the worst case when all voice flows are in the "on" state.

We show the AP-bottleneck effect with the legacy DCF MAC in Figs. 4 and 5. The frame arrival rate of an MN is constant due to voice traffic characteristics, and the arrival rate of the AP increases linearly with the number of voice connections. Because the traffic load of the AP is $N-1$ times that of an MN, collisions are more likely to occur from the viewpoint of an MN, which results in a lower service rate for an MN. However, the AP becomes unstable before MNs because of its much higher traffic load. As shown in Fig. 4, when the 30th G.729 voice connection joins in, the traffic intensity of the AP $\rho_0 > 1$, indicating the queue of the AP is no longer stable. The queue of frames at the AP will build up and all downlink flows will be eventually damaged due to the ever increasing queuing delay and buffer overflow. Therefore, with G.729 and a 10 ms sample period, at most 29 voice connections can be supported. The traffic intensities of the AP and MNs are shown in Fig. 5. It is observed that the traffic intensity of the AP is always much higher than that of an MN due to the heavier traffic load in the downlink, whichever codec is chosen.

The service rates of the AP and MNs carrying G.729 voice connections with a 10 ms interval under different MAC mechanisms are compared in Fig. 6. With the legacy DCF MAC, every MF needs to initiate a transmission and the service rates of the AP and MNs are quite low. With downlink aggregation, the AP only needs to contend once to transmit a train of frames for a single or multiple destinations. Although the AP requires a long time period to transmit the aggregated frame (which contains $N-1$ separate downlink frames), the average service time for a

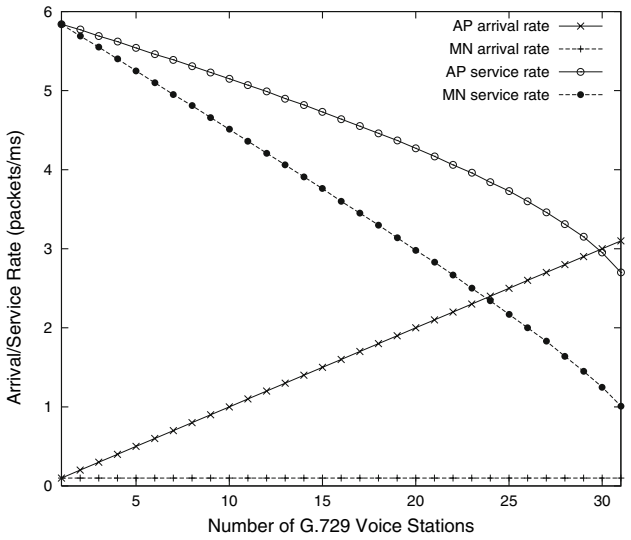


Fig. 4 Traffic arrival rate versus service rate

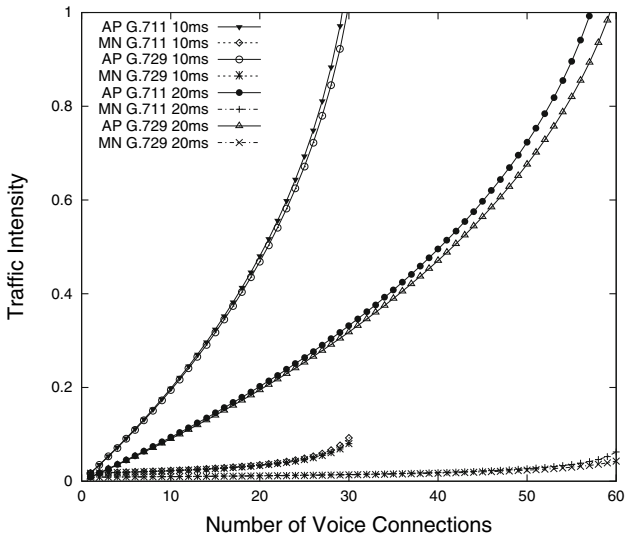


Fig. 5 Traffic intensity of the AP and MN

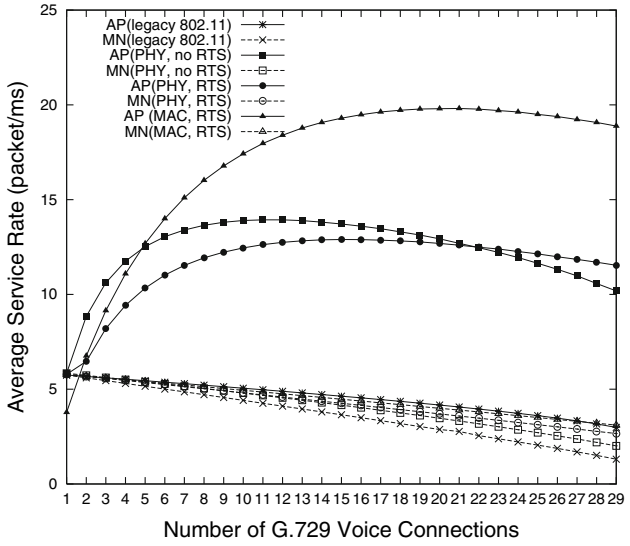


Fig. 6 Service rates of different MAC aggregation schemes

single downlink frame is significantly improved with the average service rate of $(N-1)\mu_1$ (since a train of $N-1$ frames can be served during the interval $1/\mu_1$). When the number of voice connections is small, it is more efficient to transmit multiple small G.729 packets in the basic access mode to avoid RTS/CTS overheads. When the number of connections increases, RTS/CTS outperforms basic access by reducing the collision time among large data frames. We also observe that MAC layer aggregation achieve much higher service rates of the AP and MNs than PHY level aggregation by further reducing the PHY overheads, although the scheduling information added may degrade the service rate slightly when the number of connections is quite small.

The capacity of various aggregation schemes are compared with that of legacy 802.11 in Fig. 7. It can be seen that physical aggregation in the basic access mode may not always improve the voice capacity. With a large sample period, more voice connections can be supported and more collisions may occur. Under high collisions, long frames may degrade the transmission efficiency significantly and thus reduce the voice capacity, even if the arrival rate of the AP is as low as that of an MN. When RTS/CTS is used, voice capacity can be improved by 30–40% compared to that with the legacy 802.11. Voice capacity can be further improved by around 10% with MAC-level aggregation which reduces $N-2$ physical overheads P_{OH} compared with PHY-level aggregation.

Voice capacity can be improved by 35–45% with bidirectional transmission, as shown in Fig. 8. The analysis is for CBR traffic, in which case the AP always has a frame to transmit in the reverse direction during a sample period. Since voice application is almost half-duplex, the AP may not always be able to piggyback. With the consideration of the packets transmitted during the off period, bidirectional transmission mechanism is still useful for voice applications.

5.2 Video capacity evaluation

Wireless video streaming service is another promising and demanding service in the next generation WLANs. Some video-based applications include video telephony, video conferencing, telecommuting, telemedicine, e-training, e-learning, etc. There are a large number of media platforms for video services, the majority of which employ the

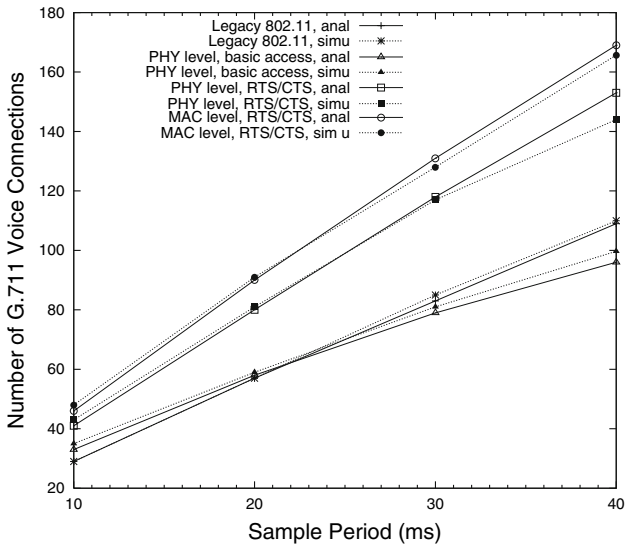


Fig. 7 Capacity comparisons of various aggregation schemes

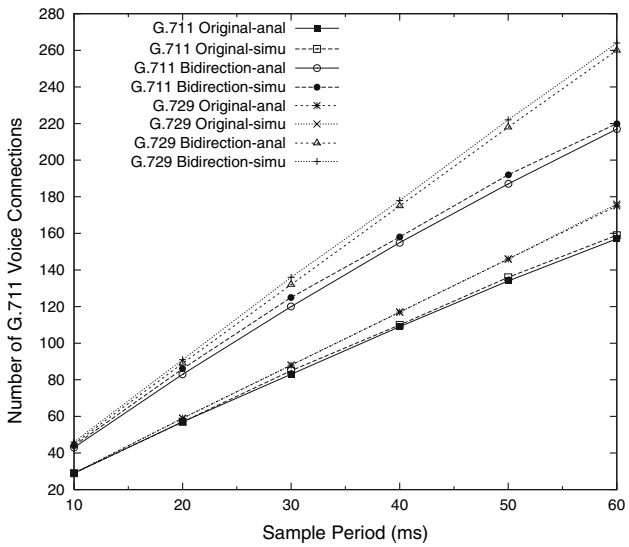


Fig. 8 Bidirectional transmission

ITU-T H.26x video standards, including H.261, H.263, and H.263+, etc. Jointly developed by the ITU-T video coding experts group (VCEG) and the ISO/IEC moving picture experts group (MPEG), H.264/MPEG-4 advanced video coding (AVC) is the latest international video coding standard that supports very high data compression. The H.264 codec has a broad range of applications that covers all forms of digital video from low rate Internet streaming applications (e.g., 64 Kbps) to broadband high definition video (HDV) applications (e.g., 240 + Mbps).

Two main objectives of H.264 video coding are to enhance the coding efficiency and improve the network adaptation. H.264 codec consists of two conceptual layers, video coding layer (VCL) and network abstraction layer (NAL) [14]. The VCL contains the signal processing functionality of the codec such as transform, quantization, motion search/compensation, and the loop filter, and outputs video slices. The NAL encapsulates the slices into NAL units (NALUs), which are suitable for transmission over packet networks. The RFC 3984 defines three RTP payload formats for H.264 codec: single NALU packet, aggregation packet and fragmentation units (FU) [15]. One NAL unit can be encapsulated in one RTP packet, fragmented over multiple RTP packets; or, multiple NAL units can be aggregated into a single RTP packet [15].

Generally, video traffic is considered as a variable bit rate (VBR) flow due to the different compression ratios in the slices and various payload formats supported in the codec. In the standard, *levels* specify the maximum frame size in terms of the total number of pixels/frame. H.264/MPEG-4 AVC defines 16 different levels, tied mainly to the picture size and frame rate [16]. Some examples for various resolution, frame rate, and maximum compressed video rate in five levels are listed in Table 3 for video capacity evaluation. At a particular level, if the picture size is smaller than the typical pictures size, then the frame rate can be higher than the typical rate. For example, the level 2 supports up to 2 Mbps video rate, with the frame rate of 30 frames per second (fps) at the frame resolution of 320×240 pixels, or with a higher frame rate of 36 fps at a lower resolution of 352×288 pixels. Higher resolution provides better image quality and higher frame rate results in a smoother motion video.

Table 3 Levels in H.264/MPEG-4 AVC [16]

Level Number	Video bit rate (bps)	Resolution & frame rate (fps)	Average frame payload (bytes)
1 b	128 k	128×96 & 15	1067
1.2	384 k	320×240 & 20	2400
1.3	768 k	352×288 & 30	3200
2	2 M	352×288 & 30	8333
		320 x 240 & 36	6944

A single video frame may be fragmented into multiple independent slices in the H.264 encoder and then encapsulated into RTP packets for transmission, depending on the maximum transmission unit (MTU) of the underlying layers.

We take level 1.3 for example. With the video rate of 768 Kbps and the frame rate of 30 fps, the required payload size is 768000/(30 * 8) = 3200 bytes. If MTU is set to be 2000 bytes, the video frame will be fragmented into two packets with the average payload of 1600 bytes in each packet, and the corresponding transmission time of each packet is $Ts = DIFS + (P_{OH} + T_{MF}) + SIFS + T_{ACK} = 34 + (24 + (1600 + 40 + 34)/216) + 16 + 24.5 = 82.25$ μs.

When supporting 2 Mbps video with the frame rate of 30 fps, the payload output from the VCL is 8333 bytes and, on average, 5 video packets are output for transmission during a sample period, which is 1000/30 = 33.33 ms.

We have obtained the video capacity of the legacy DCF MAC with the developed model. As shown in Fig. 9, the number of supported L1b video flows is quite large due to the low data rate (128 kbps) and encoding frame rate (15 fps). Low frame rate results in a longer frame interval, and thus more multiplexing gain can be achieved. The video capacity is non-decreasing with the increased MTU size. An L1b video flow requires a payload of 1067 bytes and each video frame can be encapsulated in one RTP packet for transmission when $MTU \geq 1500$ bytes, while a video frame may be fragmented and encapsulated over multiple RTP packets for transmission when a smaller MTU is used, e.g., 3 RTP packets with an average payload of 355 bytes are output for transmission when $MTU = 500$ bytes and 2 packets with an

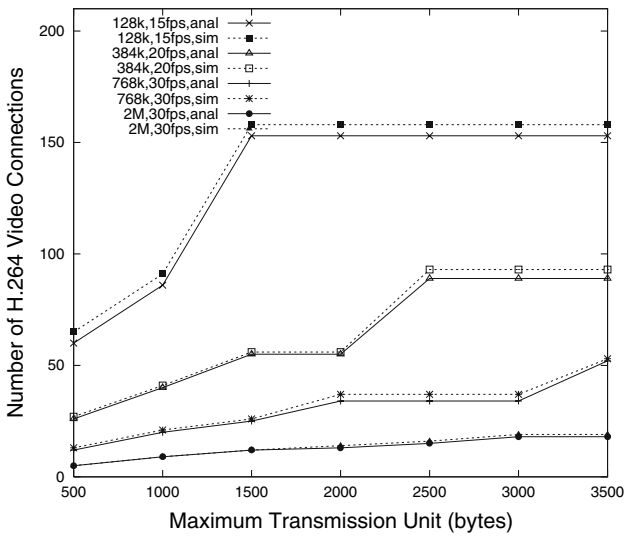


Fig. 9 Video capacity of the legacy 802.11 DCF MAC

average payload of 533 bytes when MTU = 1000 bytes. Since every packet needs to contend for transmission in the legacy DCF MAC, the MAC layer traffic arrival rate for MTU = 1500, 1000, 500 bytes are $5 \times 9 \times 10^{-6} = 0.135 \times 10^{-3}$, $15 \times 2 \times 9 \times 10^{-6} = 0.27 \times 10^{-3}$, and $15 \times 3 \times 9 \times 10^{-6} = 0.405 \times 10^{-3}$ MFs per slot, respectively. The number of output RTP packets is non-deceasing with the reducing MTU size. Thus, the MAC layer traffic arrival rate λ_i (i = 0 for AP and i = 1 for an MN) increases accordingly, which results in a lower capacity based on the AP-bottleneck constraint $(N-1)\lambda_0/\mu_0 < 1$.

Although a large number of low rate video flows can be supported with the enhanced data rate in the next generation WLAN, the video capacity for high rate video flows is still very limited. For L2 video with 2 Mbps data rate, 30 fps frame rate, and 1500 bytes MTU, a video frame is fragmented into 6 RTP packets and the MAC layer traffic arrival rate is as high as 1.62×10^{-3} MAC frames per slot. The maximum number of L2 video connections that can be supported with the legacy MAC is only 12. When MTU = 1000 bytes and 500 bytes, the traffic arrival rate increases to 2.43×10^{-3} and 4.59×10^{-3} MFs per slot, and the video capacity decreases to 9 and 5, respectively.

Video capacity under various MAC mechanisms are compared in Fig. 10. We employ a frame aggregation scheme as shown in Fig. 2(b). That is, a train of MFs that belong to one single video frame can be aggregated for one transmission. With this aggregation scheme, the MAC arrival rate is only determined by the frame rate, no matter how many fragments (or RTP packets) are output from one single video frame. However, the transmission time increases with the number of output fragments. We take L2 video with 2 Mbps data rate and 30 fps frame rate for illustration. When MTU = 1500 bytes, one video frame is fragmented over 6 RTP packets and the corresponding transmission time is $Ts = DIFS + 6(P_{OH} + T_{MF} + SIFS + T_{ACK} + SIFS) - SIFS$. It is observed that video capacity can be improved by up to 66% for L1.3 video flow (768 kbps) and 80% for L2 video (2 Mbps) with MAC level aggregation scheme in Fig. 2(b). More improvements can be achieved by further reducing the PHY, SIFS and ACK overheads using delayed ACK in Fig. 2(a), group ACK, or MAC level aggregation scheme in Fig. 2(d). When MTU = 3500 bytes, no fragmentation is required and thus the legacy MAC and the aggregation scheme achieve the same capacity performance.

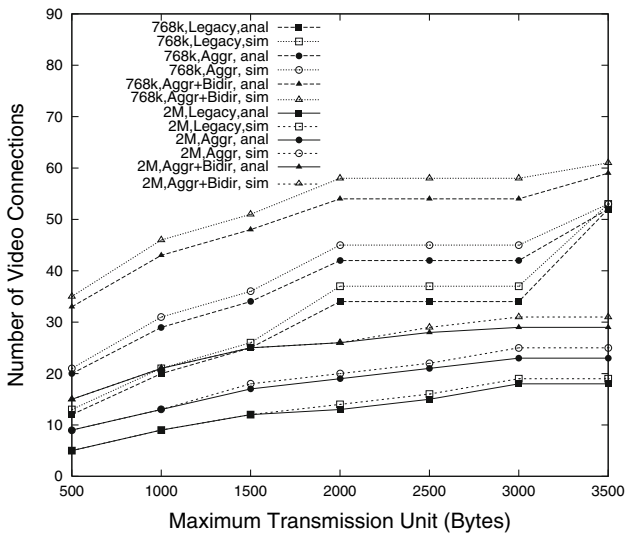


Fig. 10 Video capacity comparison under various MAC mechanisms

Downlink aggregation scheme can effectively smooth the AP bottleneck effect and efficiently improve the voice capacity, as shown in Sect. 5.1. However, it may not be appropriate for high rate video connections with much higher payload because an overly long transmission time of the aggregated downlink flow may cause serious starvation of the uplink flows. On the other hand, bidirectional transmissions can be used to eliminate the AP-bottleneck effect and improve the capacity for both video and voice services. Therefore, we combine the bidirectional transmission and aggregation schemes for video services. That is, the heavily loaded AP do not contend with MNs, but can transmit an aggregated train of MFs belonging to one single video frame upon receiving an aggregated train of uplink frames. As shown in Fig. 10, the combined scheme can improve the video capacity of legacy 802.11 MAC by 2–3 times.

6 Conclusions

We have analytically studied the capacity of an infrastructure-based WLAN supporting voice and video services, considering the various enhanced MAC mechanisms in IEEE 802.11n, i.e., frame aggregation and bidirectional transmission. Although an IEEE 802.11n WLAN can support a large number of low rate voice/video connections, the high definition video capacity is still very limited. Other MAC enhancement mechanisms, optimal aggregation parameters, and the performance of voice/video traffic under the realistic MIMO channel conditions are under investigation.

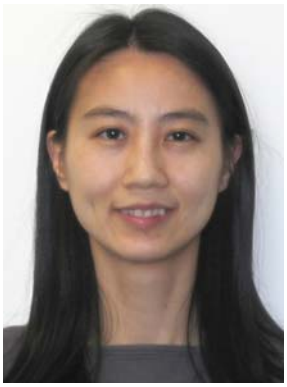
Acknowledgements This work has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant No. RGPIN7779.

References

1. Garg, S., & Kappes, M. (2003). Can I add a VoIP call? In *Proc. IEEE ICC'03* (Vol. 2, pp. 779–783).
2. Hole, D. P., & Tobagi, F. A. (2004). Capacity of an IEEE 802.11b wireless LAN supporting VoIP. In *Proc. IEEE ICC'04* (Vol. 1, pp. 196–201).

3. Cai, L. X., Shen, X., Mark, J. W., Cai, L., & Xiao, Y. (2006). Voice capacity analysis of wlan with unbalanced traffic. *IEEE Transactions on Vehicular Technology, 55*(3), 752–761.
4. Wilson, J. M. (2004). The next generation of wireless LAN emerges with 802.11n. In *Technology Intel Magazine*.
5. Bianchi, G. (2000). Performance analysis of the IEEE 802. 11. distributed coordination function. *IEEE Journal on Selected Areas in Communications, 18*(3), 535–547.
6. Tickoo, O., & Sikdar, B. (2004). A queueing model for finite load IEEE 802.11. random access MAC. In *Proc. IEEE ICC'04* (Vol. 1, pp. 175–179), Jun. 2004.
7. Xiao, Y. (2005). IEEE 802.11 performance enhancement via concatenation and piggyback mechanisms. *IEEE Transactions on Wireless Communications, 4*(5), 2182– 2192.
8. Liu, C., & Stephens, A. P. (2005). An analytic model for infrastructure wlan capacity with bidirectional frame aggregation. In *Proc. IEEE WCNC'05* (pp. 113–119).
9. Xiao, Y. (2005). IEEE 802.11n: Enhancements for higher throughput in wireless LANs. *IEEE Wireless Communication, 12*, (6), 82–91.
10. Black, U. (2001). *Ovice over IP* (2rd ed.). Prentice Hall.
11. Apostolopoulos, J. G., Tan, W., & Wee, S. J. (2002). Video streaming: concepts, algorithm, and systems. Technical report, HP. Tech. HPL-20020260.
12. Heck, A. (2003). *Introduction to maple* (3rd ed.). New York: Springer-Verlag.
13. International Telecommunication Union. G.729 annex b: A silence compression scheme for g.729 optimized for terminals conforming to recommendation v.70. In *Telecommunication Standardization section of ITU*, Nov. 1996.
14. Kim, J., Tode, H., & Murakami, K. (2006). MAC-layer support for real-time video over IEEE 802.11 DCF networks. *IEICE Transactions on Communication, E89-B*(4), 1382–1391.
15. Wenger, S., Hannuksela, M., Stockhammer, T., Westerlund, M., & Singer, D. (2005). RTP payload format for H.264 video.
16. Sullivan, G. J., Topiwala, P., & Luthra, A. (2004). The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions. In *Proc. SPIE on Applications of Digital Image Processing XXVII*.

Author Biographies

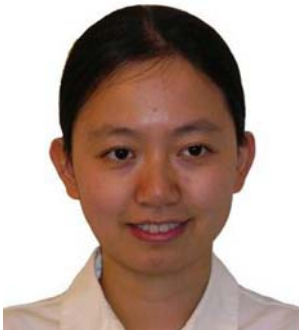


Lin X. Cai received the B.Sc. degree in Computer Science from Nanjing University of Science and Technology, Nanjing, China, in 1996 and the M.A.Sc. degree in Electrical and Computer Engineering from the University of Waterloo, Waterloo, Canada, in 2005. She is currently working toward the Ph.D. degree in the same field at the University of Waterloo. Her current research interests include network performance analysis and protocol design for multimedia applications over broadband wireless networks.

Xinhua Ling received the B.Eng. (1993) degree in Radio Engineering from Southeast University, China, the M.Eng. (2001) degree in Electrical Engineering from the National University of Singapore, Singapore, and the Ph.D. (2007) degree in Electrical and Computer Engineering from the University of Waterloo, Canada. Since May 2007, he has been a post-doctoral fellow at the University of Waterloo. His research interests are in the area of cellular, WLAN, WPAN, ad hoc networks and their internetworking, including MAC protocol design and performance analysis, and cross-layer design for Quality-of-Service support.

Xuemin (Sherman) Shen received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in Electrical Engineering. He is a Professor and the Associate Chair for Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on mobility and resource management in interconnected Wireless/Wired Networks, UWB Wireless Communications Systems, Wireless Security, and ad hoc and Sensor Networks. He is a co-author of three books, and has published more than 300 papers and book chapters in Wireless Communications and Networks, Control and Filtering. Dr. Shen serves as the Technical Program Committee Chair for IEEE Globecom'07, General Co-Chair for Chinacom'07 and QShine'06, the Founding Chair for IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as a Founding Area Editor for IEEE Transactions on Wireless Communications; Associate Editor for IEEE Transactions on Vehicular Technology; KICS/IEEE Journal of Communications and Networks; Computer Networks (Elsevier); ACM/Wireless Networks; and Wireless Communications and Mobile Computing (John Wiley), etc. He has also served as Guest Editor for IEEE JSAC, IEEE Wireless Communications, and IEEE Communications Magazine. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 from the University of Waterloo, the Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 from the Faculty of Engineering, University of Waterloo. Dr. Shen is a registered Professional Engineer of Ontario, Canada.

Jon W. Mark (LF) received the Ph.D. degree in Electrical Engineering from McMaster University in 1970. In September 1970 he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, where he is currently a Distinguished Professor Emeritus. He served as the Department Chairman during the period July 1984–June 1990. In 1996 he established the Centre for Wireless Communications (CWC) at the University of Waterloo and is currently serving as its founding Director. Dr. Mark had been on sabbatical leave at the following places: IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976–77); AT\&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant (1982–83): Laboratoire MASI, Universit\'{e} Pierre et Marie Curie, Paris France, as an Invited Professor (1990–91); and Department of Electrical Engineering, National University of Singapore, as a Visiting Professor (1994–95).He has previously worked in the areas of adaptive equalization, image and video coding, spread spectrum communications, computer communication networks, ATM switch design and traffic management. His current research interests are in broadband wireless communications, resource and mobility management, and cross domain interworking. He is a co-author of the text Wireless Communications and Networking, Prentice-Hall 2003. A Life Fellow of IEEE, Dr. Mark is the recipient of the 2000 Canadian Award for Telecommunications Research and the 2000 Award of Merit of the Education Foundation of the Federation of Chinese Canadian Professionals. He was an editor of IEEE Transactions on Communications (1983–1990), a member of the Inter-Society Steering Committee of the IEEE/ACM Transactions on Networking (1992–2003), a member of the IEEE Communications Society Awards Committee (1995–1998), an editor of Wireless Networks (1993–2004), and an associate editor of Telecommunication Systems (1994–2004).



Lin Cai (S'00-M'06) received the M.A.Sc. and Ph.D. degrees (with Outstanding Achievement in Graduate Studies Award) in Electrical and Computer Engineering from the University of Waterloo, Waterloo, Canada, in 2002 and 2005, respectively. Since July 2005, she has been an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Victoria, British Columbia, Canada. Her research interests span several areas in Wireless Communications and Networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic over wireless, mobile, ad hoc, and sensor networks. She serves as the Associate Editor for IEEE Transactions on Vehicular Technology (2007-), EURASIP Journal on Wireless Communications and Networking (2006-), and International Journal of Sensor Networks (2006-).