Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 51 (2007) 3265–3277
Computer Networks
www.elsevier.com/locate/comnet

Capacity analysis and MAC enhancement for UWB broadband wireless access networks

Lin X. Cai [a], Xuemin (Sherman) Shen [a,*], Jon W. Mark [a], Lin Cai [b]

[a] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1
[b] Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada V8W 3P6

Received 4 December 2006; accepted 14 December 2006
Available online 1 February 2007

Responsible Editor: J. Misic

Abstract

We analyze the capacity of ultra-wide band (UWB) wireless access networks supporting multimedia services by calculating the number of multimedia connections that can be supported in a UWB access network with IEEE 802.15.3 Medium Access Control (MAC) protocol, considering the overheads from different layers. We then propose how to increase the capacity by improving the MAC protocol design and parameter setting. We also quantify the effectiveness of different approaches and derive the appropriate MAC protocol parameters such as the duration of contention period and contention free period in a superframe. Simulation results are given to demonstrate the accuracy of the analysis.

Keywords: Capacity analysis; MAC enhancement; UWB

1. Introduction

Recent advances in semiconductor technology have facilitated the implementation and application of ultra-wide band (UWB) communication systems. With high data rate, low cost and low energy consumption, UWB is considered as one of the most promising wireless communications technologies. Prototypes and consumer products using UWB technologies to deliver high data rate (>100 Mbps) multimedia traffic over a short distance ($\leqslant$10 m) with very low power consumption rate have been emerging. Currently, there are two competing physical layer UWB technologies: impulse radio based direct-sequence (DS) UWB and multi-band orthogonal frequency-division multiplexing (MB OFDM)-based UWB [1,2]. Each has its own advantages and disadvantages, and both have gained supports from different industry groups. UWB devices based on these two competing technologies have been under development. However, their performance in a *networked* environment is still an open issue. To fully appreciate the potential of UWB, the capacity and other performance characteristics of UWB networks need to be well understood.

* Corresponding author. Tel.: +1 519 8884567x32691.
E-mail addresses: lcai@bbcr.uwaterloo.ca (L.X. Cai), xshen@bbcr.uwaterloo.ca (Xuemin (Sherman) Shen), jwmark@bbcr.uwaterloo.ca (J.W. Mark), cai@ece.uvic.ca (L. Cai).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2006.12.012

Although recent applications using UWB technologies are mainly with consumer electronics in personal area networks (PANs), there are increasing demands to connect these devices to the Internet to handle heterogeneous multimedia services, *e.g.*, IP Television (IPTV), video on demand (VoD) [3]. With high data rate and low energy consumption rate, UWB technologies are ideal for high-volume multimedia killer applications with stringent QoS requirements. The success of UWB multimedia services depends on how cost-effective it is to support multimedia traffic over UWB access networks with QoS guarantee. On the other hand, IEEE 802.15.3 adopts a hybrid MAC designed for high rate WPANs. To the best of our knowledge, there is little work on the performance study of hybrid MAC with respect to the configurable parameters, including the retry limit, the length of a superframe, contention and contention free periods, *etc*. Given a fixed length superframe, more devices can successfully send requests in a longer contention period, but fewer slots can be allocated for data transmissions in the contention free period, and vice versa. Therefore, appropriate lengths for contention and contention free periods are critical to IEEE 802.15.3 MAC. A simulation study of IEEE 802.15.3 MAC is presented in [4] to investigate the performance of real-time and best-effort traffic with various superframe lengths and different ACK policies. In [5], the performance of intra-piconet communications is enhanced by taking advantage of the multi-rate support in the physical layer. In [6], the relationship between the duration of contention period and the number of requesting devices is studied. The analysis in [6] is based on the result from [7] which studies the capacity of the p-persistent CSMA and may not be suitable for the traditional CSMA/CA adopted in IEEE 802.15.3. In [6,7], the channel is observed at the end of each successful transmission, and the time interval between two successful transmissions is considered regenerative. However, due to the very small size of the minimum contention window and the maximum retry limit specified in IEEE 802.15.3, which are 7 and 3, respectively, the frames are not always successfully transmitted and may be dropped after three retransmissions. The frame drop probability increases with the number of requesting devices and may not be negligible. Thus, the general assumption of no frame drop in IEEE 802.11 and the previous model in [6] does not hold in IEEE 802.15.3. Moreover, the regenerative property of each successful transmission holds when all devices continuously have packets for transmission during the contention period, but it does not hold in the case that each device only transmits at most one request per superframe, which is a special case in non-saturation scenarios. None of the existing models can be directly used for investigating this special case.

The capacity of a UWB network supporting multimedia services is defined as the number of multimedia connections that can be supported with satisfactory QoS during any one use of the UWB network. This paper analytically studies the performance of IEEE 802.15.3 MAC protocol and investigates the above defined capacity of a UWB network using the hybrid MAC. We substantiate the analysis by quantifying the voice capacity of a DS-UWB network. The analysis is extended to other multimedia applications in both DS-UWB networks and MB OFDM-UWB networks. Analytical results show that, although UWB communication networks offer high data rate, because of the overheads from different layers, the resource utilization efficiency is very low and the capacity of UWB networks employing the existing standard MAC protocol is actually limited. There are substantial margin and needs for further improvement.

To fully explore the potential of UWB networks, we first develop an analytical model to study the critical parameters employed in IEEE 802.15.3. Secondly, we propose different approaches to reduce the MAC layer overheads, and quantify the effectiveness of different approaches.

The remainder of the paper is organized as follows. We first present the system model and study the existing MAC protocols specified in the IEEE standard in Section 2. In Section 3, the capacity of a single UWB network supporting multimedia services based on the IEEE 802.15.3 MAC protocol is analyzed, and some MAC enhancement schemes are proposed to improve the capacity by reducing the overheads. The analysis is substantiated by quantifying the capacity of a DS-UWB network supporting multimedia services with various data rates in Section 4, followed by concluding remarks and future research directions in Section 5.

2. System model and IEEE 802.15.3 standard

To support high-volume multimedia applications with stringent QoS requirements, we consider an infrastructure-based UWB access network, as shown in Fig. 1. A UWB AP is located in a radio

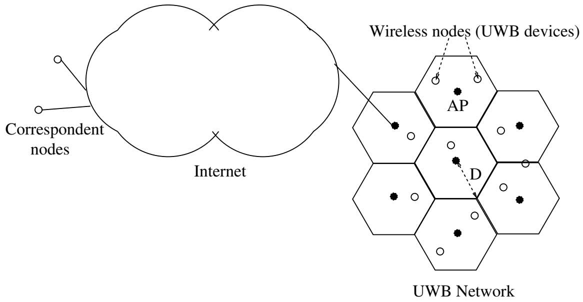


Fig. 1. Infrastructure-based UWB network.

cell, and UWB APs are connected to the Internet, either directly or via a multi-hop mesh network. The UWB devices in each cell can access the Internet through the AP.

2.1. Medium access control protocol

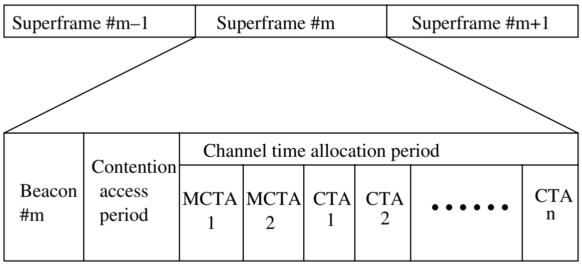
Generally, wireless MAC protocols can be classified into three categories [8]: random access, guaranteed access and hybrid access protocols. Random access protocols, such as pure Aloha, slotted Aloha, carrier sense multiple access with collision avoidance (CSMA/CA), and non/p/1-persistent CSMA, *etc*., are contention-based protocols that can operate in either infrastructure-based wireless networks or infrastructureless (ad hoc) networks. The advantage of the random access protocol is that it does not require a centralized controller nor synchronization between the devices. Asynchronous communication has been a key factor to the overwhelming success of the IEEE 802.11 wireless local area networks (WLANs). Devices are required to sense the channel before transmission to reduce collisions. According to the Federal Communications Commission (FCC) regulations, marketing and operation of UWB devices are permitted under the conditions that the mean transmission power must not exceed −41 dBm/MHz, and the peak/mean power ratio must be less than 20 dB. In other words, the power spectral density of UWB signals in any frequency band must be below the noise level. Consequently, it is very difficult and costly for UWB devices to detect whether the medium is busy or idle. On the other hand, without effective medium sensing, the impact of collisions is not negligible and may lead to poor network performance, *e.g.*, low throughput, large delay and delay jitter, especially in dense UWB networks. In addition, broadcast communication becomes more complicated in decentralized contention-based UWB networks. Without global synchronization and effective medium sensing, the intended destinations may be busy in transmitting or receiving, and the broadcast messages may easily be missed. Without reliable broadcast communications, many network management functions cannot work well, *e.g.*, the address resolution protocol (ARP), the route discovery in routing protocol, which may degrade the network performance.

Unlike random access protocols, guaranteed access are contention-free protocols in which devices access the medium in an orderly manner, usually in a round-robin fashion, and thus a certain level of quality of service can be provided. Hybrid protocols combine the best qualities of the random access and guaranteed access protocols to achieve flexibility, efficiency and QoS provisioning [8]. With the hybrid protocols, each device sends a request to the network controller indicating how much time or bandwidth is required for data transmission using a random access protocol. Based on the received requests, the controller allocates time slots and sends grants to the requesting devices indicating the start time and the duration of data transmissions. Due to the physical characteristics of UWB communications, the inefficiency of the random access protocols and the inflexibility of guaranteed access protocols, we anticipate a hybrid MAC will be deployed for UWB access networks to provide satisfactory QoS for delay-sensitive multimedia applications.

2.2. IEEE 802.15.3

IEEE 802.15.3 standard uses a hybrid MAC protocol, with random access periods for network initiation/association and resource allocation, and contention-free periods for data transmission. Several devices can autonomously form a piconet in which one of them is selected as the piconet coordinator (PNC). The PNC can collect global information about the piconet and allocate radio resources or schedule channel times to the devices in the piconet according to their requirements. Based on the scheduling, the devices can communicate in a peer-to-peer fashion. Such a semi-ad hoc setting can provide better QoS than a pure ad hoc network.

Timing in the 802.15.3 is based on the superframe, which is illustrated in Fig. 2. IEEE 802.15.3 defines three methods for communicating data between devices [9]: (a) sending asynchronous data or communicating commands in the contention access period (CAP); (b) allocating channel time



MCTA: Management Channel Time Allocations
CTA: Channel Time Allocations

Fig. 2. 802.15.3 Superframe.

for isochronous streams in the channel time allocations period (CTAP); and (c) allocating asynchronous channel time in the CTAP.

Although both commands and asynchronous data can be transmitted in the CAP, it is recommended that only commands be transmitted to minimize the length of contention period. This is desirable for reducing the protocol overheads and inefficiency resulting from potential collisions and high cost of medium sensing in UWB networks. In addition, for multimedia applications, UWB devices may need channel times on a regular basis, and they send channel time requests during the CAP to reserve isochronous channel time in the CTAP. Based on the successfully received requests from all devices, the PNC will schedule and allocate channel time in the CTAP to all devices in a time division multiple access (TDMA) manner.

When using UWB as the Internet access technology to support multimedia applications with stringent QoS requirements, we adopt the architecture of IEEE 802.15.3 in our system model. The UWB AP is the PNC of each UWB network by default.

3. Capacity analysis of 802.15.3 UWB network

An interesting feature of UWB technology, whether it be DS-UWB or MB-UWB, is that, with an efficient transceiver design, the data rate can be adjusted *proportional* to the signal-to-interference-plus-noise ratio (SINR) at the receiver [10]. This can be explained as follows. Let P denote the received signal power, r the channel capacity, and N_0 the one-sided power spectral density of white Gaussian noise plus interference.[1] According to the Shannon theory, $r = W\log_2(1+\mathrm{SINR})$ bps, where $\mathrm{SINR} = \frac{P}{N_0 W}$. Also, as pointed out by Shannon, for a system with ultra-wide bandwidth, $W \to \infty$,

$$r \approx \frac{P}{N_0}\log_2 e(\mathrm{bps}) \tag{1}$$

Accordingly, the UWB sender can always adjust its data rate (by adapting its coding) according to the arbitrary SINR to maintain the bit error rate (BER) requirement. This property holds for both DS-UWB and MB-UWB systems [1,2]. This is different from the narrow band wireless communications, such as 802.11 WLANs. For narrow band wireless communications, if the SINR falls below a certain threshold, the receiver cannot receive any meaningful data.

The unique feature of UWB communications brings many implications that are unique to UWB networks, and it has great impact on network planning and management. First, considering the large-scale propagation models of wireless communications, the average received power is proportional to $d^{-\alpha}$, where d is the distance between the UWB sender and the receiver, and α is the path loss exponent. Thus, the data rate of a UWB connection will be proportional to $d^{-\alpha}$. α depends on the environment and usually takes the value between 2 and 6. For instance, when $\alpha = 3$, the achievable data rate for a connection will increase by 7 times if the distance is reduced by half. Therefore, the capacity of a UWB network heavily depends on the network topology and user deployment. This is quite different from the capacity of narrow band wireless networks which have less scalable data rates.

We consider a piconet consisting of N UWB devices. A bi-directional multimedia connection is established between a UWB device (the mobile host) and a correspondent node through the AP, as shown in Fig. 1. Let the achievable data rate for a UWB connection with distance $d_{\max}$ be R Mbps, and the distance between the ith UWB device to the AP be d_i, where $0 \leqslant d_i \leqslant d_{\max}$ for $i = 1,2,\ldots,N$. The achievable data rate of the ith device is $R_i = (d_i/d_{\max})^{-\alpha}R$.

To support a number of multimedia connections in a UWB network, we first quantify the appropriate CAP duration to ensure that the requests for channel times can be successfully sent to PNC. Second, we calculate the number of multimedia connections that can be accommodated in the CTAP period, followed by how to reduce overheads to

[1] The Gaussian approximation holds in most of the cases. For example, as shown in [11], it holds for the Win-Scholtz UWB physical model in the presence of a large number of interferers.

improve capacity. We then substantiate the analysis by calculating how many voice and video calls can be supported in a UWB access network, considering different source coding schemes.

3.1. Contention access period

For constant rate multimedia traffic, *e.g.*, voice applications, the device may only need to send one request per call and reserve channel time for contention free data transmission in the CTAPs. Therefore, we denote the number of contending devices at the beginning of each CAP as $n \leqslant N$. As shown in Fig. 3, the process of frame service continues until all channel requests are served. Note that a served frame can be either successfully transmitted or dropped due to excessive retransmissions. Define p_k the conditional collision probability and τ_K the transmission probability of any device when there are k contending devices in the piconet. Conditioning on the number of contending devices k, we have

$$p_k = 1 - (1 - \tau_k)^{k-1}. \tag{2}$$

The exponential backoff procedure of CSMA/CA can be modeled as a truncated geometrical random variable. The average time a device takes in the backoff stage can be derived as [12]

$$E[W_k] = \sum_{i=0}^{m-1} p_k^i (1 - p_k) \sum_{j=0}^{i} \left(\frac{2^j W - 1}{2} \right) + p_k^m \sum_{j=0}^{m} \left(\frac{2^j W - 1}{2} \right), \tag{3}$$

where W is the minimum contention window and m is the retry limit. During the period of $E[W_k]$, a device makes A_k attempts to transmit a request, which can also be modeled as a truncated geometrical random variable with mean

$$E[A_k] = (1 - p_k)1 + p_k(1 - p_k)2 + \cdots + p_k^m(m + 1) = \frac{1 - p_k^{m+1}}{1 - p_k}. \tag{4}$$

Therefore, the transmission probability τ_k can be derived as

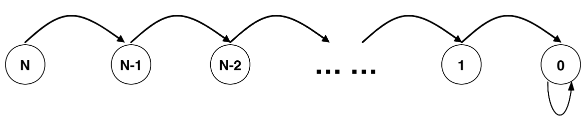


Fig. 3. Process of frame service.

$$\tau_k = \frac{E[A_k]}{E[W_k] + E[A_k]}. \tag{5}$$

Substitute (3) and (4) into (5), τ_k is a function of p_k. With (2) and (5), p_k and τ_k can be obtained. Conditioning on k contending devices in the piconet, a frame is dropped with probability p_k^{m+1} when the retry limit m is reached, and successfully transmitted with probability $1 - p_k^{m+1}$. Therefore, during the contention period, the total number of successfully transmitted requests can be approximated as

$$n_s \approx \sum_{k=1}^{n} (1 - p_k^{m+1}) \tag{6}$$

and the number of failed requests is $n_f = n - n_s$. To avoid collisions, the failed requests will not contend with others in the same CAP, but will be re-initiated in the next contention period.

For a given number of devices contending at the beginning of each CAP, the duration of the contention periods in a superframe can be considered as a renewal process. During each CAP, the number of contending devices monotonically decreases with time. To analyze the duration of the CAP, we need to track n at every time slot, which can be very complicated. To simplify the analysis, we consider the average number of contending devices over the CAP, which is $r = n/2$. We can then obtain the average conditional collision probability p and average transmission probability τ of a device. We investigate the network performance from the viewpoint of the channel status, which can be either idle or busy. The channel is idle only when there is no transmission, and sensed busy because of collisions or successful transmissions, as shown in Fig. 4. Thus, the length of CAP (T_{CAP}) is the summation of the busy periods and idle periods. Here we consider an idea duration of CAP that ends when the last request is served.

$$T_{\mathrm{CAP}} = E[N_{\mathrm{busy}}] \cdot T_{\mathrm{busy}} + E[N_{\mathrm{idle}}] \cdot T_{\mathrm{slot}}, \tag{7}$$

where $E[N_{\mathrm{busy}}]$ is the average number of busy slots, $E[N_{\mathrm{idle}}]$ is the average number of idle slots, and T_{busy} and T_{slot} are the duration of a busy slot and an idle slot, respectively. According to IEEE 802.15.3 [13], $T_{\mathrm{slot}} = 17.3\ \mu s$, and $T_{\mathrm{busy}} = T_{\mathrm{req}} + \mathrm{SIFS} + T_{\mathrm{ACK}} + \mathrm{BIFS}$, where T_{req} is the transmission time of a request. In this paper, T_{busy} is constant with fixed request frame length. To obtain the length of the CAP, we need to compute $E[N_{\mathrm{busy}}]$ and $E[N_{\mathrm{idle}}]$. Since the channel is idle only when all requesting devices are in the backoff stage, $E[N_{\mathrm{idle}}]$ is the average

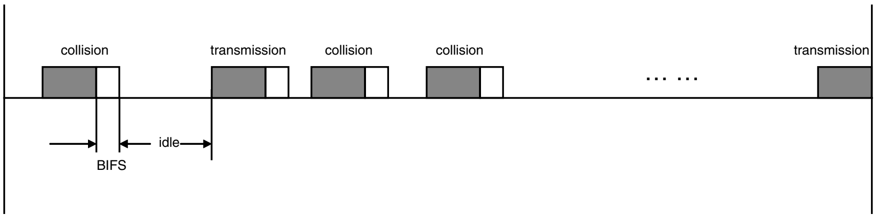


Fig. 4. Channel state during the contention access period.

backoff time the last DEV experiences, which can be approximated as

$$E[N_{\mathrm{idle}}] = \sum_{i=0}^{m-1} p^i(1-p)\sum_{j=0}^{i}\frac{2^jW-1}{2} + p^m\sum_{j=0}^{m}\frac{2^jW-1}{2}. \tag{8}$$

We calculate $E[N_{\mathrm{busy}}]$ from its probability distribution function

$$E[N_{\mathrm{busy}}] = \sum_{i=0}^{Q} i\cdot P[N_{\mathrm{busy}}=i], \tag{9}$$

where Q is the maximum number of busy slots. When the retry limit is m, n frames can be transmitted at most $(m+1)n$ times. During the CAP, n request frames can be served, but only n_{s} are successfully transmitted. Therefore, among $(m+1)n$ busy slots, n_{s} are successful transmissions and $(m+1)n - n_{\mathrm{s}}$ are collisions. With most collisions caused by two simultaneous transmissions, the maximum number of busy slots is $Q = [(m+1)n + n_{\mathrm{s}}]/2$.

Given the channel is busy, implying at least one device is transmitting in the given slot, two events may occur: (1) a collision resulting from multiple simultaneous transmissions; (2) a successful transmission when only one device transmits in that slot. Thus, the probability of a collision and a successful transmission can be derived as

$$p_{\mathrm{c}} = \frac{1-(1-\tau)^r - r\tau(1-\tau)^r}{1-(1-\tau)^r}, \tag{10}$$

$$p_{\mathrm{s}} = \frac{r\tau(1-\tau)^r}{1-(1-\tau)^r}. \tag{11}$$

The number of busy slots can be considered as a binomial random variable. In other words, in i busy slots, n_{s} slots are due to successful transmissions and the remaining $i - n_{\mathrm{s}}$ slots are due to collisions.

$$P[N_{\mathrm{busy}}=i] = \binom{i}{n_s}(p_s)^{n_{\mathrm{s}}}(p_c)^{i-n_{\mathrm{s}}} \tag{12}$$

Substitute (8)–(12) into (7), we can obtain the CAP duration in which n requests can be served but only n_s requests are transmitted successfully.

3.2. Channel time allocation period

As shown in Fig. 5, a short interframe space (SIFS) time is required to ensure sufficient turnaround time between transmissions. Let T_{CTA} be the time unit for a channel time allocation unit, $T_{\mathrm{CTA}} = k(T_{\mathrm{frame}} + \mathrm{SIFS} + T_{\mathrm{ACK}} + \mathrm{SIFS})$, where $k = 1$ without link layer fragmentation and $k > 1$ in other cases. Denote p_{l} the payload of a packet. The frame payload, including p_{l} and RTP/UDP/IP headers ($\mathrm{RUI_h}$), will be transmitted at a data rate of R_i. The PHY header ($\mathrm{PHY_h}$), MAC header ($\mathrm{MAC_h}$), header check sequence (HCS) and frame check sequence (FCS) will be transmitted at a lower basic rate of R_0. The preamble is used for clock/carrier acquisition and receiver training. The length of the preamble should be long enough for synchronization between the AP and all UWB devices in the network, *i.e.*, the preamble time T_{a} of the ith device is determined by d_i. Therefore, T_{frame} and T_{ACK} are obtained as

$$T_{\mathrm{frame}} = T_{\mathrm{a}} + (p_{\mathrm{l}} + \mathrm{RUI_h})/R_i + (\mathrm{PHY_h} + \mathrm{MAC_h} + \mathrm{HCS} + \mathrm{FCS})/R_0,$$
$$T_{\mathrm{ACK}} = T_{\mathrm{a}} + (\mathrm{PHY_h} + \mathrm{MAC_h} + \mathrm{HCS})/R_0.$$

In a TDMA system, guard times (T_{g}) are required to keep transmissions in adjacent CTAs from colliding. Let the AP and the N devices exchange a pair of packets in one superframe. Then the duration of the superframe, T_{sf}, is

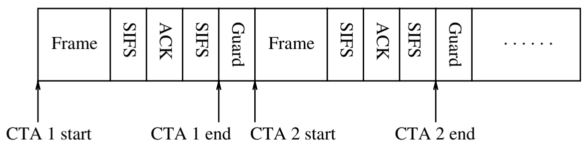


Fig. 5. Channel time allocation with TDMA.

$$T_{\mathrm{sf}} = T_{\mathrm{beacon}} + T_{\mathrm{CAP}} + (2N+1)T_{\mathrm{g}} + 2NT_{\mathrm{CTA}}. \quad (13)$$

Assume all multimedia connections have a constant packet inter-arrival time. If T_{sf} is less than the packet inter-arrival time, all N connections can be supported in the UWB network. Since $d_i \leqslant d_{\max}$ for all connections, the lower bound of the number of multimedia connections being supported, $N_{\min}$, can be obtained as

$$N_{\min} = \frac{T_{\mathrm{sf}} - T_{\mathrm{beacon}} - T_{\mathrm{CAP}} - T_{\mathrm{g}}}{2T_{\mathrm{g}} + 4k\mathrm{SIFS} + 2kT_{\mathrm{frame}} + 2kT_{\mathrm{ACK}}}. \quad (14)$$

3.3. Reducing overheads for capacity improvement

To improve the network capacity, it is necessary to reduce as many overheads as possible. The overheads include header overheads of different layers, preamble or acquisition time for frame synchronization, guard times, interframe space SIFS, and ACK, *etc*. Header overheads, such as $\mathrm{RUI_h}$, $\mathrm{PHY_h}$ and $\mathrm{MAC_h}$ are usually unavoidable, but can be compressed at the expense of additional processing time. Acquisition time is hardware and distance dependent in UWB networks [14]. Some policies, such as *piggyback* and *No-ACK*, can effectively reduce the overheads of SIFS, ACK and guard times.

Piggyback: As shown in Fig. 6, the UWB AP only needs to allocate one CTA for a two-way multimedia connection by allowing the AP and the UWB device to exchange their frames. Therefore, N CTAs are required for N multimedia connections while $2N$ CTAs are needed when traditional TDMA is used. The duration of the superframe is

$$T_{\mathrm{sf}} = T_{\mathrm{beacon}} + T_{\mathrm{CAP}} + (N+1)T_{\mathrm{g}} + NT_{\mathrm{CTA}},$$

where $T_{\mathrm{CTA}} = 2kT_{\mathrm{frame}} + (2k+1)\mathrm{SIFS} + T_{\mathrm{ACK}}$. In addition, $N_{\min}$ is obtained as

$$N_{\min} = \frac{T_{\mathrm{sf}} - T_{\mathrm{beacon}} - T_{\mathrm{CAP}} - T_{\mathrm{g}}}{T_{\mathrm{g}} + (2k+1)\mathrm{SIFS} + 2kT_{\mathrm{frame}} + T_{\mathrm{ACK}}}. \quad (15)$$

No-ACK: With error-resilient source coding schemes, many multimedia applications can tolerate a certain degree of packet loss, and thus *No-ACK*

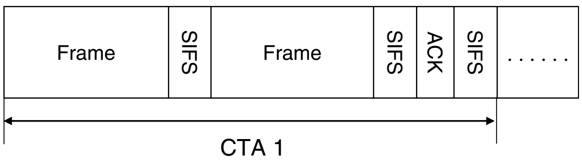


Fig. 6. Channel time allocation with Piggyback.

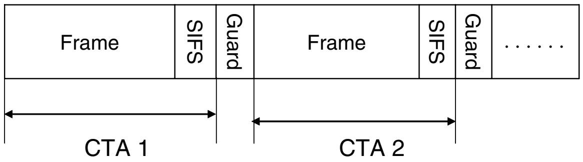


Fig. 7. Channel time allocation with *No-ACK* and separate CTA.

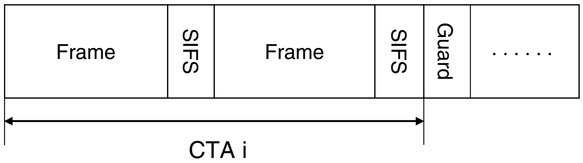


Fig. 8. Channel time allocation with *No-ACK* and combined CTA.

policy can be used to reduce the ACK overhead. In this case, T_{CTA} is reduced to $T_{\mathrm{CTA}} = kT_{\mathrm{frame}} + k$ SIFS, as shown in Fig. 7. Accordingly, $N_{\min}$ is obtained as

$$N_{\min} = \frac{T_{\mathrm{sf}} - T_{\mathrm{beacon}} - T_{\mathrm{CAP}} - T_{\mathrm{g}}}{2T_{\mathrm{g}} + 2k\mathrm{SIFS} + 2kT_{\mathrm{frame}}}. \quad (16)$$

The overhead of guard times can be further reduced by allowing the sender and receiver to exchange data in one CTA. With the combined CTA, the duration of the superframe is

$$T_{\mathrm{sf}} = T_{\mathrm{beacon}} + T_{\mathrm{CAP}} + (N+1)T_{\mathrm{g}} + NT_{\mathrm{CTA}},$$

where $T_{\mathrm{CTA}} = 2kT_{\mathrm{frame}} + 2k$ SIFS, as shown in Fig. 8. Compared with separate CTA, the number of guard time is reduced from $2N+1$ to $N+1$. Therefore, $N_{\min}$ is obtained as

$$N_{\min} = \frac{T_{\mathrm{sf}} - T_{\mathrm{beacon}} - T_{\mathrm{CAP}} - T_{\mathrm{g}}}{T_{\mathrm{g}} + 2k\mathrm{SIFS} + 2kT_{\mathrm{frame}}}. \quad (17)$$

No-ACK can effectively reduce the ACK overhead at the cost of reliability because it disables the link layer recovery. Therefore, *No-ACK* is not desired when channel error rate is high. On the other hand, *piggyback* is highly recommended when the devices in the UWB networks carry symmetric traffic, such as two-way voice communications. With dynamically asymmetric traffic, it is difficult for the AP to efficiently allocate combined CTAs for the pair of sender and receiver.

4. Voice and video capacity estimation

To substantiate the analysis, we first calculate the voice capacity of a DS-UWB network in this subsection as an example, since voice capacity of other wireless networks (*e.g.*, cellular, WLAN) has been heavily studied and it can be used for direct comparison. We then calculate the maximum number of high data rate multimedia connections being supported in a UWB network with different data rates. To maximize the capacity, we obtain the required minimum duration of CAP. Extensive simulations are performed to validate the analysis using NS-2 [15].

Let a UWB AP support a large number of VoIP calls with a comparatively large range, 10 m. The system parameters for DS-UWB are listed in Table 1. DS-UWB supports two independent bands of operation. The lower band occupies the spectrum from 3.1 GHz to 4.85 GHz and the upper band occupies the spectrum from 6.2 GHz to 9.7 GHz. R_0 is 28 Mbps for the lower band, and 55 Mbps for the upper band. We choose the basic rate $R_0 = 28$ Mbps and data rate $R = 110$ Mbps to calculate the voice capacity in this paper. There are three preambles defined in [1]: (1) short preamble: 5 μs in length that requires a high SINR with low channel dispersion – it is most suitable for short range links (<3 m); (2) nominal preamble, 15 μs in length that requires a nominal SINR with a nominal channel – it is the default preamble choice; and (3) long preamble, 30 μs in length that is used for a poor SINR and/or highly dispersive channel. Long preamble is intended for extended range applications. The Beacon of each superframe contains a variable number of information elements (IEs), 4 bytes FCS, and 21 bytes synchronization parameters [9]. The beacon length is approximately 200 bytes plus $34n$ bytes for CTA IEs and CTA status IEs. Guard times are calculated based on the worst-case drift in a superframe and the maximum allowed number of lost beacons g [9], $T_g = g * (\text{Clock accuracy}/10^6) * T_{sf}$, where $g = 10$ and the clock accuracy is ±25 ppm.

Table 1
System parameters

Data rate R	110 Mbps
Basic rate R_0	28 Mbps
Preamble T_a	5 or 15 or 30 μs [1]
PHY header	2 Bytes [1]
MAC header	10 Bytes [9]
HCS	2 Bytes [1]
FCS	4 Bytes [1]
SIFS	10 μs [9]
IP/UDP/RTP header	40 Bytes

Table 2 lists the main attributes of some frequently used voice codecs with different packetization intervals. 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. If the packetization interval is 10 ms, corresponding to a rate of 100 packets per second, the payload size is $64{,}000/(100 * 8) = 80$ bytes. If the packetization interval is increased to 20 ms, corresponding to a rate of 50 packets per second, the payload size is increased to 160 bytes. G.723, G.729, and internet low bitrate codec (iLBC) are popular codecs used by VoIP applications. They have lower bit rates at a cost of more 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. iLBC is developed for robust voice communications that can achieve a graceful degradation of voice quality with severe packet losses, so it is chosen by many Internet soft-phone applications, *e.g.*, Skype. iLBC has a codec bit rate of 13.3 Kbps for a 30 ms packetization interval and 15.2 kbps for a 20 ms interval.

Table 2
Frequently used voice codecs

Voice codec		G.711	G.723a	G.729	iLBC
Codec bit rate (kbps)		(64)	(5.3/6.3)	(8)	(15.2/13.3)
Sample period (ms)	Arrival Rate (frames/sec)	Payload (Byte)	Payload (Byte)	Payload (Byte)	Payload (Byte)
10	100	80		10	
20	50	160		20	38
30	33.33	240	20/24	30	50
40	25	320		40	
50	20	400		50	
60	16.67	480	40/48	60	

We first determine the proper duration of CAP. At the beginning of a superframe, each contending station sends a request with the starting time randomly chosen over the minimum contention window, *i.e.*, [0, 7]. The required CAP duration is obtained that n requests are served, either successfully received or dropped. The channel request frames are transmitted at the basic rate of 28 Mbps and the data frames are transmitted at a data rate of 110 Mbps. Fig. 9 shows the number of failed requests for a given number of contending devices and retry limit m. It is observed that the increase rate of the failed transmissions goes higher with a larger n and smaller m. In IEEE 802.11, the minimum contention window CW_{min} and m are relatively large ($CW_{min} = 32$ and $m = 7$), and most frames will be eventually transmitted successfully during a long period. But in IEEE 802.15.3, due to the much smaller minimum contention window ($CW_{min} = 8$) and the retry limit ($m \leqslant 3$), the collision probability could be very high when n is large. The request frames are more likely to be dropped and thus fewer requests can be successfully transmitted. If a request is dropped in the CAP, the device needs to re-initiate the request in the following superframe, which degrades the network performance due to the waste of network resources caused by retransmissions, and also affects user-perceived QoS with longer delay. The duration of the CAP for a given n and m is shown in Fig. 10. Note that not all served requests can be successfully transmitted. For $m \leqslant 1$, most request frames are dropped due to excessive transmissions for a large n, as shown in Fig. 9, and it is no use to enlarge CAP in this case. It is recommended that a large m be used to improve the number of successful requests when n is large. Although the AP cannot change the CAP duration in every superframe, it is necessary for the AP to determine the CAP duration from time to time, based on the varying number of contending devices in the network. For real time multimedia connections that require channel time in a regular basis, we expect that each connection can last for a certain period, and there are only a limited number of new connections that contend in the CAP. In the simulation, we set the retry limit to $m = 3$, the maximum value defined in IEEE 802.15.3. We choose $T_{CAP} = 2$ ms when $T_{sf} = 10$ ms, which is suitable to serve 12 contending devices. The TCAP increases by 0.5 ms when the superframe length increases by 10 ms to accommodate more connections. When $T_{sf} = 60$ ms, $T_{CAP} = 4.5$ ms, which is suitable for serving 20 contending devices.

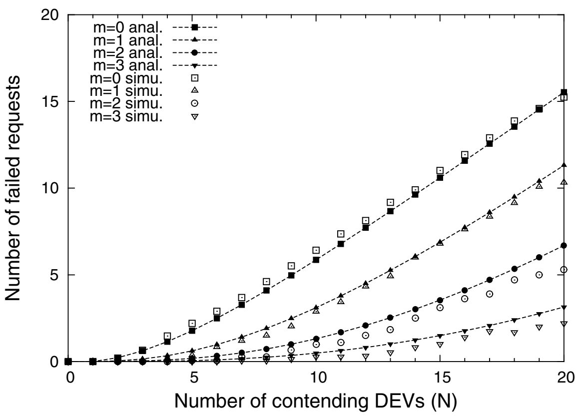


Fig. 9. Number of failed requests.

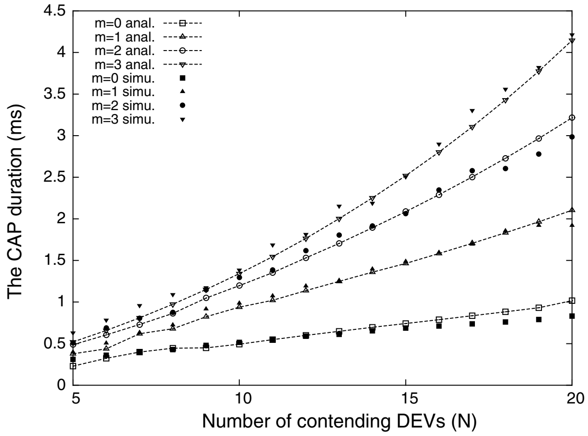


Fig. 10. The CAP duration (ms).

Without considering the silence suppression, from (14), a DS-UWB AP can support around 37 G.711 voice calls or 39 G.729 voice calls in the lower band, with a 30 μs preamble and a 10 ms packetization interval. With a 20 ms packetization interval, the DS-UWB AP can support about 76 G.711 voice calls, or 84 G.729 voice calls. The voice capacity increases with the packetization interval, as shown in Fig. 11. The increase ratio of the number of G.729 voice calls is higher than that of G.711 due to the small voice payload. The voice payload of G.711 is eight times of that of G.729, but the voice capacity of G.729 is only slightly higher than that of G.711 (less than 30%). High compression codec cannot significantly improve the voice capacity of the UWB network. We also observe that the voice capacity doubles with a 20 ms packetization

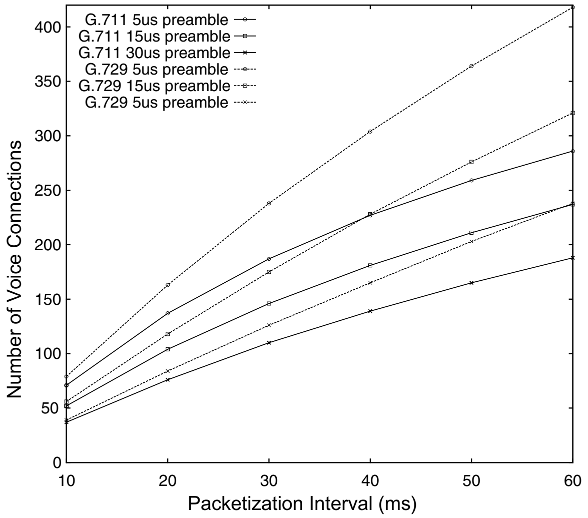


Fig. 11. Voice capacity with different preamble length.

interval. In addition, if the preamble is reduced from 30 μs to 5 μs, 137 G.711 or 163 G.729 voice calls can be supported. This is because the overheads from the PHY layer to the transport layer protocols, especially the acquisition time T_a, are the dominant factors affecting the network capacity. The DS-UWB AP can support up to 286 G.711 and 418 G.729 voice calls when the preamble is 5 μs, and the packetization interval With such a high voice capacity, a UWB AP can support voice calls with extremely high user density.

The number of voice calls supported in a UWB network for different voice codecs is listed in Table 3. For low rate voice applications over UWB networks, it is recommended to use low compression codecs with low complexity and high voice quality. The voice capacity of UWB networks will surpass that of all existing wireless technologies by a large margin. UWB has great potential to support other high data rate multimedia services that are otherwise impossible with existing narrow band communications technologies. On the other hand, without the protocol overheads, a 110 Mbps channel should be able to support 110,000/64 ≈ 1718 G.711 voice calls with 64 kbps codec rate. The actual capacity of a UWB network is far below that upper bound, which reflects the inefficiency of the protocol. Some multimedia applications, *e.g.*, professional phase alternative line (PAL) using MPEG-2 video compression, usually require high data rate in the order of Mbps. 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 rang 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) [16]. We list some examples of high data rate video connections and calculate their capacity in Table 4. The superframe length is 30 ms, and the CAP duration is 3 ms. With link layer fragmentation, one video packet corresponds to several link layer frames transmitted in a CTA. The average payload of a frame is 1250 bytes. Therefore, for a 2 Mbps video flow, the number of link layer frames transmitted in a 30 ms interval is $2 \times 10^6 \times 30 \times 10^{-3}/(1250 \times 8) = 6$. It is observed that the number of high rate video connections being supported is very limited due to the high bandwidth requirement and protocol overheads. Only 32 one Mbps two-way video connections can be supported, and the

Table 4
Number of video connections supported in a UWB network with different data rates

Preamble (μs)	Data rate				
	1 Mbps	2 Mbps	3 Mbps	4 Mbps	5 Mbps
5	32	16	11	8	6
15	28	14	9	7	5
30	24	12	8	6	4

Table 3
Number of VoIP connections supported in a UWB network with different preamble and different voice codecs

Audio (ms)	G.711			G.729			G.723			iLBC		
	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs	5 μs	15 μs	30 μs
10	71	52	37	79	56	39						
20	137	104	76	163	118	84				157	115	82
30	187	146	110	238	175	126	238/237	176/175	126/126	229	171	124
40	227	181	139	304	228	165						
50	259	211	165	364	276	203						
60	286	237	188	418	321	238	433/431	330/329	243/242			

video capacity decreases proportionally when the rates of video connection increases. Compare to voice connections, where the payload size is quite small and no frame fragmentation is needed, the fragmentation overheads of video traffic (*e.g.*, SIFSs between frames, header overheads in each frame, *etc.*) are quite large. The number of supported connections is even less if high definition television (HDTV) is considered, which requires high bandwidth from tens of Mbps (using efficient compression algorithms) to up to 1.5 Gbps (without compression). Thus, there remain considerable issues that need further improvement.

The MAC enhancement schemes introduced in Section 3.3, *piggyback*, *No-ACK* with separate CTA or combined CTA, can be used to effectively improve the network capacity. The capacity improvement with different policies is shown in Fig. 12. For G.729 with 5 μs preamble time, 79 voice calls with a 10 ms packetization interval can be supported in a UWB network based on 802.15.3; voice capacity increases to 100 with *piggyback*, 132 with *No-ACK* and separate CTA, and 159 with *No-ACK* and combined CTA. With a larger packetization interval, more voice calls can be supported. Voice capacity increases by using the *No-ACK* policy which reduces the SIFS and ACK overheads. Guard times are almost negligible when the number of CTAs is small and the duration of the frame transmission is relatively large, but becomes significant when the number of required CTAs increases. When *piggyback* and *No-ACK* with combined CTA are used, the voice capacity can be greatly improved.

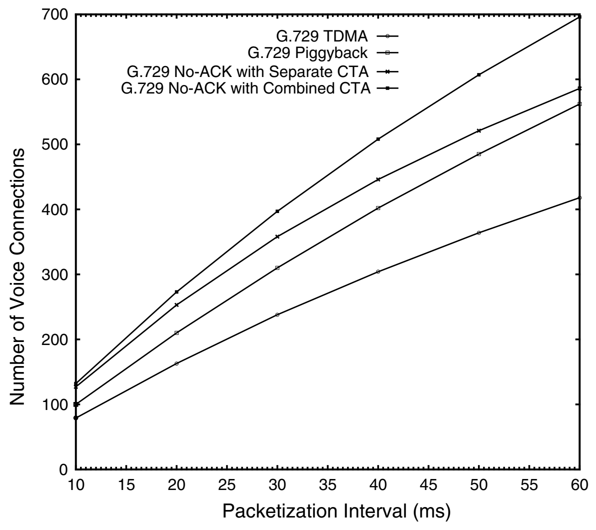


Fig. 12. Voice capacity with different policies.

The numbers of supported voice or video connections under different policies are compared in Table 5. The preamble is 30 μs and superframe length is 10 ms. It is observed that with *piggyback* and *No-ACK* policies, the number of high data rate video connections being supported can only be improved by around 15–35%, while the voice capacity can be improved by 60–85%. This is because the overheads of SIFS and ACK become relatively small compared with the large video payload. Therefore, piggyback is more efficient for small payload and highly recommended for two way symmetric, low data rate voice applications. With a limited number of high data rate connections, the guard times are negligible and there is no capacity difference under separate and combined CTA policies.

5. Conclusions

In this paper, we have analyzed the capacity of a UWB access network supporting multimedia services based on the IEEE 802.15.3 architecture, considering the overheads from different layers. We have further investigated how to increase the capacity by reducing the MAC overheads, and quantified how effective these approaches are. The capacity of UWB networks supporting variable bit rate video traffic with multiplexing gain is currently under investigation.

Table 5
Number of connections supported in a UWB network using different policies

MAC policies	Data rate						
	8 kbps	64 kbps	1 Mbps	2 Mbps	3 Mbps	4 Mbps	5 Mbps
TDMA	39	37	20	10	7	5	4
Piggyback	51	48	23	12	8	6	5
No-ACK with separate CTA	70	64	27	13	9	7	5
No-ACK with combined CTA	72	66	27	13	9	7	5

Acknowledgements

This work has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Strategic Grant # STPGP 257682, and a research grant from Bell University Laboratories (BUL) under the sponsorship of Bell Canada.

References

[1] IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a. DS-UWB physical layer proposal. IEEE P802.15-04/0137r4, January 2005.
[2] IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a. Multiband OFDM physical layer proposal, September 2004.
[3] Y. Chu, A. Ganz, MAC protocols for multimedia support in UWB-based wireless networks, in: Proceedings of the International Workshop on Broadband Wireless Multimedia (Broad-Wim'04).
[4] K.-W. Chin, D. Lowe, Simulation study of the IEEE 802.15.3 MAC, in: Proceedings of the Australian Telecommunications and Network Applications Conference (ATNAC), December 2004.
[5] Z. Yin, V.C.M. Leung, Connection data rate optimization of IEEE 802.15.3 scatternets with multi-rate carriers, International Journal of Sensor Networks 3(6) (2008), in press.
[6] Y.-H. Tseng, H.-K. Wu, G.-H. Chen, Maximum traffic scheduling and capacity analysis for IEEE 802.15.3 high data rate MAC protocol, in: Proceedings of the VTC'03, 3 October 2003, pp. 1678–1682.
[7] F. Cali, M. Conti, E. Gregori, Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit, in: IEEE/ACM Transactions on Networking.
[8] A.C.V. Gummalla, J.O. Limb, Wireless medium access control protocols, IEEE Commun. Surveys (2000).
[9] IEEE standard part 15.3: Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs). IEEE Std 802.15.3-2003, September 2003.
[10] S. Verdu, Spectral efficiency in the wideband regime, IEEE Trans. Inf. Theory 48 (6) (2002) 1319–1343.
[11] M. Win, R. Scholtz, Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications, IEEE Trans. Commun. 48 (4) (2000) 679–689.
[12] L.X. Cai, X. Shen, J.W. Mark, L. Cai, Y. Xiao, Voice capacity analysis of wlan with unbalanced traffic, IEEE Trans. Veh. Technol. 55 (2006) 752–761.
[13] IEEE 802.15.3 WG. IEEE std 802.15.3™ – 2003: Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs), September 2003.
[14] X. Shen, W. Zhuang, H. Jiang, J. Cai, Medium access control in ultra-wideband wireless networks, IEEE Trans. Veh. Technol. 54 (5) (2005) 1663–1677.
[15] NS-2. Network Simulator (2.27). Available in <http://www.isi.edu/nsnam/ns/>.
[16] G.J. Sullivan, P. Topiwala, A. Luthra, 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, Aug. 2004.

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 a 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 access networks.

Xuemin (Sherman) Shen (M 1997 – SM 2002) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since October 1993, where he is a professor and the Associate Chair for Graduate Studies. His research focuses on mobility and resource management in interconnected wireless/wireline networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a co-author of two books, and has published more than 200 papers and book chapters in wireless communications and networks, control, and filtering. He was Technical Co-Chair for the IEEE GLOBECOM 2003, ISPAN 2004, QShine 2005, IEEE Broadnets 2005, and WirelessCom 2005, and is Special Track Chair of the 2005 IFIP Networking Conference. He serves as Associate Editor for IEEE Transactions on Wireless Communications; IEEE Transactions on Vehicular Technology; Computer Networks; ACM/Wireless Networks; Wireless Communications and Mobile Computing (Wiley); and International Journal Computer and applications. He has also served as Guest Editor for IEEE JSAC, IEEE Wireless Communications, and IEEE Communications Magazine. He received the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, for demonstrated excellence of scientific and academic contributions, and the Distinguished Performance Award in 2002 from the Faculty of Engineering, University of Waterloo, for outstanding contributions in teaching, scholarship, and service.

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 to 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. He had been on sabbatical leave at the following places: IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976–1977); AT&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant (1982–1983): Laboratoire MASI, Universitíäe Pierre et Marie Curie, Paris France, as an Invited Professor (1990–1991); and Department of Electrical Engineering, National University of Singapore, as a Visiting Professor (1994–1995). 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, he 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 (M 2006 – S 2000) 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, Victoria, 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. Her research work has been published in prestigious journals and conferences (*e.g.*, IEEE Transactions on Mobile Computing, IEEE Transactions on Wireless Communications, IEEE Transactions on Multimedia, IEEE Transactions on Vehicular Technology, ACM MobiCom, IEEE Infocom). She serves as the Associate Editor for EURASIP Journal on Wireless Communications and Networking, and International Journal of Sensor Networks (IJSNet).