Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Resource management for video streaming in ad hoc networks

Ruonan Zhang [a], Lin Cai [b], Jianping Pan [c], Xuemin (Sherman) Shen [d,*]

[a] School of Electronics & Information, Northwestern Polytechnical University, Xi'an, Shaanxi, China
[b] Dept. of Electrical & Computer Engineering, University of Victoria, Victoria, BC, Canada
[c] Dept. of Computer Science, University of Victoria, Victoria, BC, Canada
[d] Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Medium access control
Resource management
Performance modeling
Multimedia services

ABSTRACT

Video streaming over wireless links is a challenging issue due to the stringent Quality-of-Service (QoS) requirements of video traffic, the limited wireless channel bandwidth and the broadcast nature of wireless medium. As contention-based or reservation-based (*i.e.*, contention-free) medium access control (MAC) protocols in existing wireless link-layer standards cannot efficiently support multimedia applications such as video streaming, a hybrid approach has been proposed, which uses both contention and reservation-based channel access mechanisms to transmit packets for each video source. Using this content-aware resource management approach, each video source reserves well below its peak data rate, and uses contention-based media access to transmit the remainder of the packets. In this paper, we first propose two conflict avoidance strategies and two buffering architectures for video streaming over ad hoc networks. Considering the interactions of reservation and contention, we develop the analytical model for the saturated traffic case and then extend it to derive tight performance bounds for the unsaturated case. Using the MAC protocols specified in the WiMedia ECMA-368 standard as an example, extensive simulations have been conducted to validate the analysis. Real video traces have been used to examine the video streaming performance. The analytical and simulation results demonstrate the effectiveness and efficiency of the hybrid resource management approach, and also reveal the impact of the conflict avoidance strategy and buffering design on the video performance.

1. Introduction

Video streaming over wireless networks is of increasing demand, and it becomes possible thanks to the emerging high-speed, short-range wireless communication technologies such as IEEE 802.11n, Ultra-WideBand (UWB) and Milli Meter Wave (MMW). However, supporting wireless video streaming is a challenging issue due to the stringent Quality-of-Service (QoS) requirements of video traffic, the limited wireless channel bandwidth and the broadcast nature of wireless medium. How to utilize the premium wireless resources and support the QoS for video streaming is a challenging issue. Most of the existing wireless standards define either contention-based or contention-free medium access control (MAC) protocols, and some newer ones support both of them in each media access cycle, such as the WiMedia standard (*e.g.*, EMCA-368) for UWB networks.

Contention-based MAC protocols have been widely used for providing wireless data services, such as the IEEE 802.11 Distributed Coordination Function (DCF) protocol in Wireless Local Area Networks (WLANs). But this approach suffers from an intrinsic weakness for supporting video traffic: when there are multiple sources in the system, even though the sum of their average data rates is below the channel capacity, excessive queueing delay and even packet losses are still possible due to collisions and

* Corresponding author. Tel.: +1 519 8884567x32691.
E-mail address: xshen@bbcr.uwaterloo.ca (Xuemin (Sherman) Shen).

1570-8705/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2010.08.012

backoff procedures, which may lead to a considerable degradation of video quality.

Reservation-based (*i.e.*, contention-free) protocols are often preferred for video streaming, which guarantee the channel access opportunity to satisfy the delay requirement. However, due to the burstiness of video traffic, if reservations are made at the peak data rate to minimize queuing delay, the channel utilization is low; if made at the average data rate, considerable queuing delay and even packet losses can occur during traffic bursts. Therefore, neither reservation nor contention alone is an effective approach for supporting video streaming in wireless networks.

In [1], a hybrid approach for video streaming over wireless networks has been proposed, which uses both contention and reservation-based channel access for transmitting packets from each video flow, by reserving well below the peak data rate and handling traffic bursts using contention-based channel access. By properly selecting the number of time slots reserved for each video stream, it is possible to efficiently utilize the reserved slots and reduce the competition level during the contention periods. To use both reservation and contention access in ad hoc networks, how to define the conflict avoidance strategy in the presence of reservation periods and how to design the buffering architecture are open issues.

In this paper, we first propose two conflict avoidance strategies for reservation and contention interleaved wireless systems. When a station obtains a Transmission Opportunity (TXOP) but there is insufficient time to complete the frame transaction before the arrival of the next reserved period, the station may have two conflict avoidance strategies: the *hold-on strategy* by which the station just holds on its frame and transmits immediately after the reserved period; or the *backoff strategy* by which the station invokes a new backoff procedure. Which strategy has better performance with different traffic conditions is not obvious and requires in-depth investigation.

We then propose the single-buffer and dual-buffer design for utilizing both reservation and contention periods. With dual-buffer, the video packets being transmitted using contention or reservation-based channel access are separated and stored in two buffers. As shown in the paper later, the novel dual-buffer can considerably improve the efficiency (*i.e.*, increase the utilization of reserved slots and reduce the contention level) in most scenarios. The dual-buffer is also simple to implement.

Furthermore, we develop analytical models considering the interactions of reservation and contention periods. The reservation-based channel access is much more deterministic compared to the contention-based access. On the other hand, the behavior of the contention-based channel access is significantly affected by the reservation, which is essentially different from the traditional contention-only MAC (like the IEEE 802.11 MAC). Therefore, we focus on the performance of the contention-based access. Using the *mean value analysis* method, we first establish the model for the saturated traffic case and then extend it to derive tight performance bounds for the unsaturated case. The analysis is of low computational complexity so it can be used for on-line admission control and optimizing the per-flow reservation.

The rest of the paper is structured as follows. We briefly overview the MAC protocols supporting wireless video streaming and the related work in Section 2. System models and the single- and dual-buffer design are presented in Section 3. In Section 4, we develop analytical models for both saturated and unsaturated station cases, followed by performance evaluation of video streaming through extensive simulation in Section 5. The concluding remarks and future research issues are summarized in Section 6.

2. Related work

2.1. MAC standards

Advanced video coding schemes such as MPEG-4 AVC (H.264) achieve a considerably lower data rate and better picture quality when compared with previous schemes (e.g., MPEG-2). With higher compression efficiency, these coding schemes also introduce a much higher peak-to-average data rate ratio and the applications become very sensitive to packet losses or excessive delay. Consequently, it is very challenging to support high-quality video streaming over wireless links.

MAC is critical for wireless networks due to the broadcast nature of wireless medium. The majority of WLANs products in the market (based on IEEE 802.11 standards) only support contention-based MAC. Even with the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) protocol which supports service differentiation, there is no QoS guarantee for video traffic. The IEEE 802.15a/c standards target on supporting high data rate applications with short transmission ranges. The IEEE 802.15 MAC protocols follow a superframe structure with both contention and reservation periods in each superframe. But there is a need for a centralized piconet controller, which may be complicated and with high overheads as mobile devices may join and leave the piconet at any time, and switch between active and idle state frequently.

The WiMedia ECMA-368 [2] standard is defined for UWB-based WPANs, and it has gained more popularity recently. Its MAC protocol also has a superframe structure and supports both reservation and contention-based medium access. In specific, the channel time is partitioned into superframes and each superframe is divided into 256 Media Access Slots (MASs). Inside one superframe, two MAC protocols are supported: contention-based Prioritized Contention Access (PCA) and reservation-based Distributed Reservation Protocol (DRP). DRP can negotiate and reserve MAS slots for exclusive access in a distributed manner without any centralized controller, and the non-DRP slots are available for PCA, which is similar to IEEE 802.11e EDCA. In a superframe, there can be multiple, interleaved DRP and PCA clusters. Thus, using both reservation and contention access for supporting video streaming is ready to be adopted in wireless ad hoc networks based on the WiMedia ECMA-368 standard.

2.2. Related work

Video streaming over wireless networks has attracted a lot of attention. Ref. [3] proposed a retry limit adaptation framework for video traffic, considering traffic characteristics and network conditions. In [4], cross-layer approaches were proposed to map application-layer video characteristics such as frame types to network and MAC-layer parameters such as transmission rate, retry limit, and IEEE 802.11e priorities. In [5], the performance of video streaming over a WiMedia UWB testbed was evaluated, which revealed the tradeoff between various reservation patterns.

There has been a rich literature on the performance study of contention-based MAC (*e.g.*, IEEE 802.11 MAC) with different techniques, notably among which are the Markov-chain models [6], the equilibrium point analysis [7] and the mean value analysis [8,9]. The QoS requirements of multimedia applications also prompted the study of prioritized contention-based MACs, *e.g.*, [10] for IEEE 802.11e and [11] for PCA.

The reservation-based MAC has also attracted much attention recently due to its superiority for multimedia traffic. The centralized TDMA protocol and its variants have been extensively studied in the literature. Liu [12] and Zhang [13] studied the queuing behavior and the delay performance of ECMA-368 DRP, considering the arbitrary reservation pattern and indoor UWB fading channels. Daneshi [14] investigated the allocation algorithms to make optimal reservations for multimedia streaming flows over DRP.

Most existing work studied the performance of contention or reservation-based MAC separately, but their performance when they co-exist is less explored. In [1], we proposed the hybrid MAC approach and developed a simple analytical model. Simulation results have validated the efficacy of the hybrid approach. In this paper, we propose different conflict avoidance strategies and single- or dual-buffer architecture to efficiently utilize both reservation and contention access periods and develop system performance bounds and mean values for unsaturated and saturated networks, respectively.[1] In addition, we examine the system performance using real video traces.

3. Protocol description and system model

3.1. Contention/Reservation interleaved MAC

We consider a general contention/reservation interleaved MAC approach. Suppose in a network, there are N transmitting stations, which can hear each other and thus in one collision domain. Each station has one video flow to be sent towards its destination. These stations can reserve channel access time with the coordination of a central controller (*e.g.*, following the IEEE 802.15.3 standard) or by themselves in a distributed manner (*e.g.*, following the WiMedia ECMA-368 standard). Considering fairness, each station reserves a duration of T_R in a *scheduling cycle*, denoted by T_Y, as shown in Fig. 1. The time interval between two consecutive reserved periods, denoted by T_C, is available for contention-based channel access. Each multimedia flow can obtain guaranteed channel access periodically in their reserved time slots, and can also compete with each other during the contention access periods.

During the reserved time period, the reservation owner can transmit the backlogged packets back-to-back and then receive a Block Acknowledgment (B-ACK) at the end of the burst. Such burst transmission can reduce protocol overheads and increase the channel utilization. Following the B-ACK, there are one Short Inter-Frame Space (SIFS) and a Guard Time (GT), in order to separate the reservation and contention periods and ensure the switch of operation mode of the involved stations. As the GT must be finished within the reserved period, the transmission burst size is limited accordingly.

3.2. Contention-based channel access with conflict avoidance

Since the channel access based on reservation is deterministic, we focus on the contention-based access, which follows the CSMA/CA and exponential backoff schemes, similar to the IEEE 802.11 DCF protocol. When the backoff counter is decreased to zero, the station obtains the TXOP, during which a single data frame and an Immediate ACK (I-ACK) may be exchanged. The duration of the TXOP is denoted as ϕ and $\phi = T_{DATA} + SIFS + T_{ACK}$, where T_{DATA} and T_{ACK} are the transmission time of a data frame and the I-ACK frame, respectively.

The contention-based channel access in hybrid MAC is different from the traditional contention-only MAC in the following aspects, due to the presence of the reserved channel time. First, a station shall freeze the backoff counter once the medium becomes busy (during frame transaction) or unavailable for contention (during reserved periods), and it shall sense the channel to be available for contention and idle for an Arbitration Inter-Frame Space (AIFS) before starting to decrement the backoff counter. Thus, the reservation can affect the performance of the contention access by enlarging the backoff slot and waiting time considerably. Second, when a station obtains TXOP, it needs to ensure that the whole frame transaction including I-ACK should finish at least one SIFS plus one GT before the beginning of the incoming reservation period. The time interval of $T_F = \phi + SIFS + GT$ is thus called *conflict time*. If a station obtains TXOP during T_F, to avoid conflict with the next reserved time slot, it should perform conflict avoidance procedure according to one of the two strategies: the *hold-on strategy* by which the station just holds on its frame and transmits immediately after the reserved period plus an AIFS; or the *backoff strategy* by which the station invokes another stage of backoff procedure, *i.e.*, selecting a new backoff counter.

As shown in Fig. 1, the stations can perform backoff inside the time interval between the end of the previous reservation period plus AIFS and the beginning of the next reservation, which is denoted as $T_B (= T_C - AIFS)$. T_B is further divided into T_A and T_V. The frame transactions initiated in the current contention period are completed during T_A (called *access time*). Note that because a transaction may finish inside T_F before the next reservation period, the time period left in T_C, *i.e.* T_V (called *vulnerable time*), is smaller than T_F. The channel is always idle during T_V. Consequently, all active stations will decrement their backoff counters,

[1] Part of the results reported in this paper has been presented in IEEE ICC'10, May 2010.

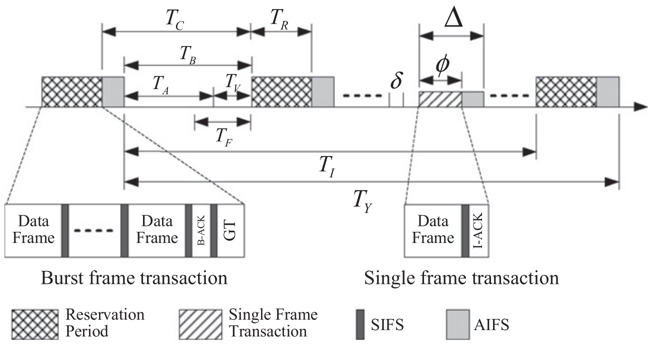
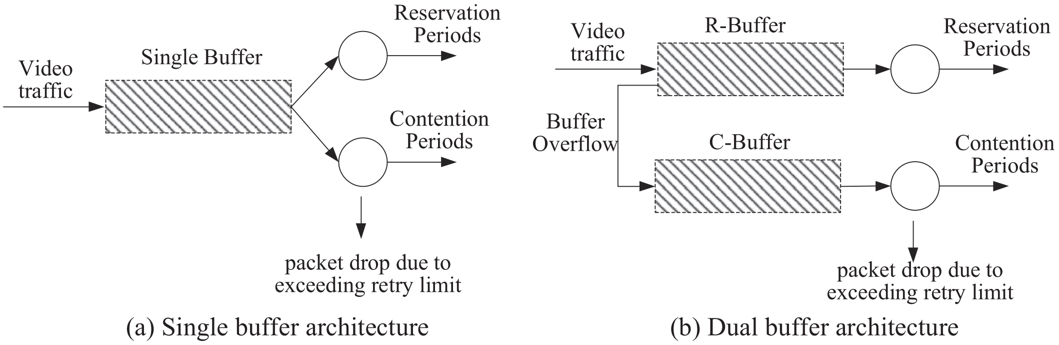


Fig. 1. System model of the proposed MAC.



(a) Single buffer architecture

(b) Dual buffer architecture

Fig. 2. Buffering architectures.

and several of them may obtain TXOP during T_V. If using the hold-on conflict avoidance strategy, their transmissions after the reserved period will collide. If using the backoff strategy, the stations obtaining TXOP during T_V will initiate another round of backoff, so their collision probability is minimized.

3.3. Single-buffer and dual-buffer

The hybrid MAC approach can be accompanied with different buffering structures. We investigate two buffering architectures for video streaming. The single-buffer and the dual-buffer architectures are illustrated in Fig. 2a and b, respectively. For the single-buffer one, there is only one logical buffer for the video traffic. When the reserved period becomes available for this video flow, packets accumulated in the buffer will be transmitted as many as possible. After its reserved time, the remaining packets can be transmitted by competing with other flows during contention periods.

One major issue with the single-buffer architecture is that, due to the asynchronous nature between video traffic and reservation schedule, it is possible that when a reserved time begins, there is insufficient video packets to send, and thus the reserved period is under-utilized and the contention periods become more crowded.

Using a dual-buffer architecture can solve the problem. As shown in Fig. 2b, the size of the R-Buffer is set according to the maximum number of packets that can be transmitted in a reservation period. Video traffic will first fill up the R-Buffer, which will be transmitted in the upcoming reservation period. When the R-Buffer is full, packets are stored in the C-Buffer, and they will contend for the channel during contention periods.[2] In this case, the utilization of the reservation periods can be maximized, so the collisions during contention periods can be minimized. Note that, dual-buffer will introduce packet reordering, as packets in the R-Buffer have to wait for the next reservation period, and those in the C-Buffer may experience random queueing delay. We need to quantify the system performance to adjust the system parameters carefully to ensure the QoS for video traffic.

4. Performance analysis

In this work, we study the performance of the contention-based channel access using the *mean value analysis*, which evaluates the average values of the system variables, such as station transmission probability, collision probability, and frame service time, without considering the details of the stochastic backoff process. Ref. [15] first studied DCF for saturated stations using this technique. In [16,8], the approach was applied to EDCA for saturated and

[2] Note that if transmission errors happen during the contention periods, the packets that need to be retransmitted can be queued in the C-Buffer as well. In this paper, we ignore the transmission errors to simplify the analysis.

unsaturated stations cases, respectively. This approach is based on the equilibrium conditions of the network: (1) for every single node, the mean values of the system variables must satisfy the given relationship among them, and (2) all of the nodes are statistically equivalent thanks to the fair nature of the contention-based MAC, which results in the same statistics of each node. Thus, an equation set can be developed and solved.

In this section, the analytical model for the saturated station case is presented first, and then we extend it to study the unsaturated case. All MAC frames are assumed to have the same length and transmission error is assumed negligible.

4.1. PCA with saturated stations

4.1.1. Transmission probability

Define the time for a sensing station to decrement its backoff counter as one generic time slot, which may be of different duration according to different channel status as discussed later. The key performance parameters considered are the collision probability and the service time of a frame, where the service time includes both the backoff and transmission time slots.

Let $E[B]$ and $E[R]$ denote the average number of backoff slots and transmission slots experienced by a frame. Given that a station is busy, *i.e.*, the station is performing backoff or transmitting, the probability to transmit in a slot equals

$$\tau = \frac{E[R]}{E[B] + E[R]}. \tag{1}$$

Let P denote the collision probability for the tagged station when it sends a frame. After collision, the station tries to retransmit the frame. The number of transmission trials of a frame follows a truncated geometric distribution with the success probability of $1 - P$. $E[B]$ and $E[R]$ can be obtained by

$$\begin{cases} E[B] = \sum\limits_{k=1}^{K} \left(\frac{CW_k}{2} P^{k-1}\right), \\ E[R] = \sum\limits_{k=1}^{K} P^{k-1}, \end{cases}$$

where CW_k and $\frac{CW_k}{2}$ are the contention window size and the average number of backoff slots in the k-th backoff stage, respectively, and K is the retry limit.

4.1.2. PCA access period

During T_A, a time slot can have three states. First, a slot may be idle if no station transmits. The duration of an idle slot is δ, and within T_A, the probability of a slot to be idle is a_A.

Second, a slot may contain a frame transaction, either successful or with collision, with probability of b_A. The duration of this type of slots is $\Delta = \phi + AIFS$.

Third, if a frame transaction begins during the interval of $[T_F, \Delta]$ before a reservation period (as defined in ECMA-368, $\Delta > T_F$), then after the frame transaction, the reserved time period will begin within an AIFS. Thus, all the contention stations have no chance to perform backoff after this frame transaction. We can regard the reservation period together with the frame transaction as one time slot with probability of b_{AD} and the duration of Δ'. We approximate that, if such a scenario happens, the beginning time of the frame transaction is uniformly distributed inside $[T_F, \Delta]$ before the reservation. Thus, the average duration of this type of transaction slots is $\Delta' = \frac{\Delta + T_F}{2}$. Further, given that a frame transaction occurs within T_A, the probability for the transaction begins during $[T_F, \Delta]$ is $\frac{\Delta - T_F}{T_A}$. In this scenario, we also have $T_V = 0$.

The probabilities for a slot being in the three states are given by

$$\begin{cases} a_A = (1 - \tau)^N, \\ b_{AD} = (1 - a_A)\frac{\Delta - T_F}{T_A}, \\ b_A = 1 - a_A - b_{AD}. \end{cases} \tag{2}$$

Thus, in T_A, the average slot duration is

$$S_A = a_A\delta + b_A\Delta + b_{AD}\Delta'. \tag{3}$$

If a transaction begins inside the time interval $[T_F, T_F + \Delta]$ before the beginning of a reservation period, the transaction will end inside the conflict time, which makes the vulnerable time T_V smaller than T_F on average. If a transaction begins inside the interval of $[T_F, T_F + \Delta]$, we approximate that its starting time is uniformly distributed, so the average vulnerable time is $\frac{T_F}{2}$. However, if no transmission occurs inside the interval, the vulnerable time is T_F. The number of idle slots in the interval $[T_F, T_F + \Delta]$ is $\Gamma_\Delta = \frac{\Delta}{\delta}$. So, the probability of no transmission initiated during the interval is $a_A^{\Gamma_\Delta}$. The average length of the vulnerable time is

$$T_V = \left(1 - a_A^{\Gamma_\Delta}\right)\frac{T_F}{2} + a_A^{\Gamma_\Delta} T_F = \left(1 + a_A^{\Gamma_\Delta}\right)\frac{T_F}{2}. \tag{4}$$

Then, the average length of T_A can be obtained by $T_A = T_B - T_V$, and the average number of slots in T_A is $\Gamma_A = \frac{T_A}{S_A}$.

4.1.3. Vulnerable time and reservation period

Because the contention stations cannot transmit during T_V and the following reserved period, we consider these two periods together. A slot inside T_V has two states. The upcoming reservation period arrives during the last slot, which is thereby enlarged by T_R plus AIFS. We approximate that the arrival time of the following reserved period is uniformly distributed inside the last idle slot in T_V. Thus, the average duration of the last idle slot is

$$\delta_D = \frac{\delta}{2} + T_R + AIFS. \tag{5}$$

All the other slots within T_V are idle slots with the duration of δ. The average duration of T_V is given in (4) and hence the average number of slots can be estimated by $\Gamma_V = \frac{T_V}{\delta}$. Finally, given a slot within the vulnerable time, the probability to have the duration of δ (*i.e.*, not the last one) is $g = \frac{\Gamma_V - 1}{\Gamma_V}$.

4.1.4. Generic channel slot

Considering the total number of generic slots during T_B, the probability for a generic slot to be in the vulnerable time is

$$h = \frac{\Gamma_V}{\Gamma_A + \Gamma_V}. \tag{6}$$

Please cite this article in press as: R. Zhang et al., Resource management for video streaming in ad hoc networks, Ad Hoc Netw. (2010), doi:10.1016/j.adhoc.2010.08.012

In summary, from the viewpoint of the entire channel time, the generic channel slots have four states: idle slots within T_A and T_V periods with the duration of δ and probability of a; frame transaction slots within T_A with the duration of Δ and probability of b; after the frame transaction, if the incoming reservation period starts within an AIFS, the transaction slot is combined with the reservation period and they are regarded as one slot, with the duration of $\Delta_D = \Delta' + T_R + AIFS$ and probability of b_D; and the last idle slot inside T_V is combined with the following reservation period as one slot, with the duration of δ_D and probability of a_D. The state probabilities are

$$\begin{cases} a = hg + (1-h)a_A, \\ b = (1-h)b_A, \\ b_D = (1-h)b_{AD}, \\ a_D = h(1-g). \end{cases} \tag{7}$$

Thus, the average length of a generic slot is

$$S = a\delta + a_D\delta_D + b\Delta + b_D\Delta_D. \tag{8}$$

4.1.5. Collision probability

Given that the tagged contention station transmits in a slot outside the conflict time, collision will happen if any other stations also transmit in the same slot. For saturated stations, the probability to transmit in a slot is τ, given in (1). On the other hand, if the TXOP is obtained within the vulnerable time, the station has to defer the transmission according to the conflict avoidance strategy.

Using the backoff strategy, the conflict can be regarded as a *virtual collision* with the reserved time periods, which results in a new backoff stage. Because the probability for a slot being in the vulnerable time is h given in (6), the overall collision probability (including virtual collision) is

$$P = 1 - (1-h)(1-\tau)^{N-1}. \tag{9}$$

By solving the Eqs. (1)–(9) with the numerical method, we can obtain the mean values of the system variables, such as P, τ, etc.

For the hold-on strategy, the virtual collision will not result in a new backoff stage, but the stations will experience collisions when more than one of them obtain TXOP within the vulnerable time. In this case, the collision probability is

$$P = 1 - (1-h)(1-\tau)^{N-1} - h(1-\tau)^{(N-1)\Gamma_V}. \tag{10}$$

Similarly, we can use the numerical method to obtain P and τ for the hold-on strategy. Due to the space limitation, we focus on the backoff strategy in the remainder of this section.

4.1.6. Average frame service time and station throughput

Because the transmission of a frame experiences $E[B] + E[R]$ generic slots, the service time is

$$\Phi = (E[B] + E[R])S. \tag{11}$$

Since only one frame is (re-)transmitted during a frame service time on average, the rate to send a new frame is $\frac{1}{\Phi}$. When all the transmissions (up to the specified retry limit, K) of a frame have failed, the frame will be discarded. Therefore, the per-station throughput (bps) can be obtained as

$$\Psi = \frac{L_P}{\Phi}(1 - P^K), \tag{12}$$

where L_P is the payload size of a frame in bits.

4.2. PCA with unsaturated stations

Different from the saturated station case, an unsaturated station only contends for channel access when its buffer is non-empty. The key parameter is the probability that the station is busy in a selected generic slot, called *busy probability*.

4.2.1. Correlated channel access

In the previous work on contention-based MAC, it is assumed that each station is busy *independently* with certain probability. However, the assumption for the unsaturated stations to be busy independently is not valid, as the channel sensing effort from a station may be blocked by the ongoing transaction or the reservation period. As a result, the probability for another station to be busy given that the tagged station is busy (*i.e.*, the conditional busy probability), is higher than the probability for another station to be busy in a randomly selected slot (*i.e.*, unconditional busy probability).

In this section, we extend the mean value analysis to the hybrid MAC with unsaturated stations. Because it is very complicated, if not impossible, to obtain the accurate conditional busy probability of another station given that the tagged station is busy, we develop the lower and upper bounds of the system performance. As shown by the numerical results, these two bounds are tight and they *converge* when the stations become saturated.

4.2.2. Lower bound

Because the probability for another station to be busy conditioned on that the tagged station is busy is larger than the unconditional busy probability, we can obtain the lower-bound of collision probability by using the unconditional busy probability in a randomly selected slot. The symbols with the superscript $'$ represent the lower-bound parameters.

First, the unconditional busy probability for a station in a randomly selected slot is

$$\rho' = \min\left\{\frac{(R' + B')S'}{\mu},\ 1\right\} \tag{13}$$

where μ is the average arrival interval of the frames and S' is the average length of a generic slot for the entire channel time. In (13), $\frac{\mu}{S'}$ gives the average number of generic slots between two consecutive arrivals. Given the station is busy, the transmission probability is τ', same as the saturated case in (1). Thus, an unsaturated station transmits in a generic slot inside the contention period with probability $\rho'\tau'$. The probabilities of the three states of a slot inside T_A are changed to

$$\begin{cases} a'_A = (1-\rho'\tau')^N, \\ b'_{AD} = (1-a'_A)\frac{\Delta - T_F}{T_A}, \\ b'_A = 1 - a'_A - b'_{AD}. \end{cases} \tag{14}$$

In addition, the collision probability is changed to

$$P\prime = 1-(1-h')(1-\rho'\tau')^{N-1}. \tag{15}$$

Following the procedure in Section 4.1, we can obtain the unsaturated versions of the other equations, which have the same form as that for the saturated case. Combined together with (13)–(15), the mean values of all the system variables can be solved. Finally the frame service time Φ' is obtained similarly to that in (11).

Different from the saturated case, the throughput for unsaturated stations depends on the incoming traffic load. Here we ignore the limit of MAC buffer size and the packet drop is caused by exceeding the retry limit only. Since the lower-bound of collision probability is P', the upper-bound of station throughput is

$$\Psi' = \frac{L_P}{\mu}\left[1-(P')^K\right]. \tag{16}$$

4.2.3. Upper bound

In a network with all stations unsaturated, there must be some time periods that all stations are non-busy. In other words, the probability that there is no station busy should be larger than zero. Therefore, the upper-bound of the collision probability is obtained by assuming that at any time moment, there is at least one station being busy. The symbols with the superscript $''$ represent the upper-bound parameters.

Under this hypothesis, the probabilities of the three states of a slot inside T_A becomes

$$\begin{cases} a''_A = (1-\tau'')(1-\rho''\tau'')^{N-1}, \\ b''_{AD} = (1-a''_A)\frac{\Delta - T_F}{T_A}, \\ b''_A = 1 - a''_A - b''_{AD}. \end{cases} \tag{17}$$

Except the new version of (17), all the other equations have the same form as that for the lower-bound case in the previous subsection. For example, the station busy probability ρ'' and the collision probability P'' can be calculated similarly using (13) and (15), respectively. By solving this new equation set, we can obtain the mean values of the system variables.

Note that, the average length of a slot conditioned on that a station is busy is larger than the average length of a randomly selected slot. Therefore, S'' (obtained by assuming there is at least one station busy) is the upper-bound of the average length of a generic slot in a network with all stations unsaturated. Finally, the lower-bound of the station throughput can be obtained in a similar way.

4.2.4. Asymptotic property

An unsaturated station can be driven into saturation when its traffic load increases, or when more channel time is reserved, which increases the occurrence of conflict (virtual collision) and also the backoff slot duration, such that the frame service time becomes larger than the arrival interval.

When the stations in the network become saturated, the approximation for the lower-bound model (*i.e.*, the busy probability of another station conditioned on that the tagged station is busy is equal to the unconditional busy probability) becomes accurate because both probabilities approach 1. On the other hand, with all stations becoming saturated, the hypothesis for the upper-bound model that there is always at least one node being busy, is satisfied. Therefore, the lower-bound and upper-bound both converge to the analytical results of the saturated case. In other words, the convergence of the two bounds indicates that the contention stations become saturated.

5. Performance evaluation

To verify the accuracy of the analysis, extensive simulations have been conducted using a discrete event-driven simulator. All the numerical results reported here are obtained based on the WiMedia ECMA-368 standard [2]. As shown in Fig. 1, T_R is the average length of the DRP reserved periods (for the analysis of PCA, the beacon period can be regarded as a DRP period), while T_C is the average length of the contention-based PCA periods between two consecutive DRP periods. If there are D DRP periods in one superframe, we have $T_C = \frac{T_{SF}}{D} - T_R$, where T_{SF} = 65,536 μs is the superframe length.

The PHY and MAC parameters defined in ECMA-368 standard are listed in Table 1. The PHY layer data rate is 480 Mbps and the payload size of a frame is 1000 bytes. Since we simulate the transfer of video streams, each packet is encapsulated in IP/UDP/RTP. The overhead of IP, UDP and RTP are 20, 8 and 12 bytes, respectively.

Note that, in the following figures, the collision probability of backoff strategy includes both the real collisions between the PCA transactions and the virtual collisions with the DRP periods.

5.1. Saturated stations, backoff strategy

For the network setting, we assume that there are N PCA and N DRP stations. Each DRP station reserves M DRP periods and each period contains one MAS (256 μs). Thus there are totally $D = NM$ DRP periods evenly distributed inside one superframe and T_R = 256 μs.

For the saturated case, each station is backlogged with packets to send. Fig. 3 shows the overall collision probability and the average frame service time, with N = 4 and N = 6 flows, respectively, using the backoff conflict avoidance strategy. We can see that there is a good agreement between the analytical and simulation results.

The subfigures indicate how the performance of contention-based access changes when more channel time is

Table 1
Parameters used for performance evaluation.

AIFS	28 μs	SIFS	10 μs	GT	12 μs
T_{DATA}	31.9 μs	T_{ACK}	13.1 μs	δ	9 μs
T_{DRP}	256 μs	K	7	CW_1	7

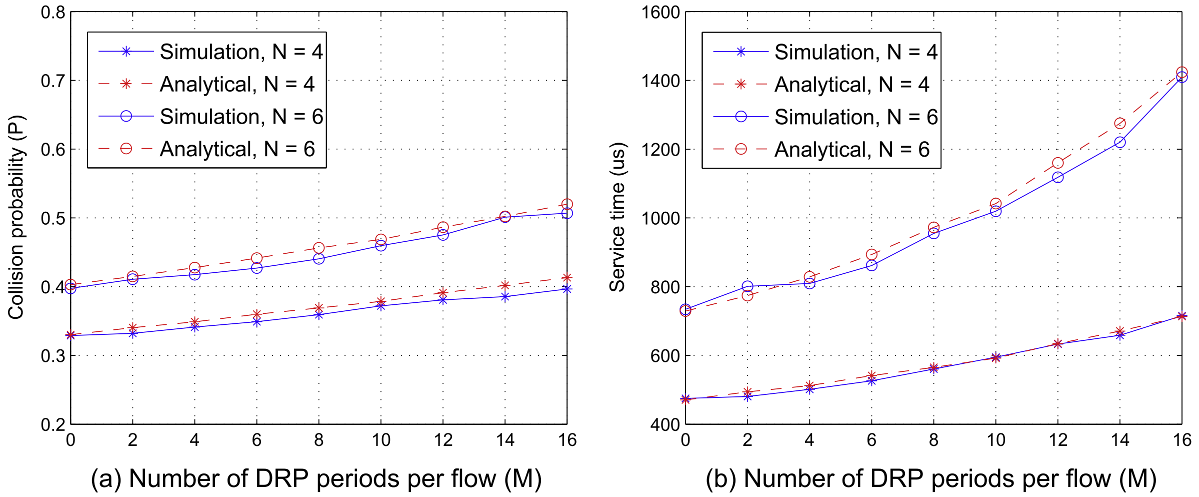


Fig. 3. Performance *vs.* M for saturated PCA stations.

reserved (as M increases). In Fig. 3a, we observe a slow increase of the overall collision probability. With larger M, the DRP periods occur more frequently, which increases the virtual collision probability. However, due to virtual collisions, the average CW is larger than that of PCA only MAC. Then the transmission probability in a given time slot is reduced and the real collisions among the PCA stations are actually reduced. Therefore, increasing M does not have significant impact on the overall collision probability.

However, as shown in Fig. 3b, the service time increases fast w.r.t. M. This is expected because the presence of DRP periods enlarges the backoff slots considerably.

5.2. Unsaturated stations, backoff strategy

For the unsaturated case, we assume that the packet arrival events at each PCA station follow a Poisson process with the average inter-arrival time of $\mu = 1000$ µs. The reservations from the N DRP stations are the same as above. Fig. 4 shows the collision probability and average service time, respectively, with $N = 6$ flows. The performances of saturated case are also plotted for comparison.

When M increases, the collision probability and service time increase accordingly, similar to the saturated case. Note that when $M \leqslant 10$, the service time is smaller than the arrival interval $\mu = 1000$ µs (as shown in Fig. 4b) and thus the stations are unsaturated. We can see that the analytical models can give valid lower and upper bounds. However, when $M \geqslant 12$, the service time becomes larger than μ, which indicates that the stations have become saturated. Both bounds converge to the results with saturated station, as expected due to the asymptotic property given in Section 4.2. Also we can see that the convergence of the lower and upper-bound models correctly predicts the transition from unsaturated to saturated status (when μ is smaller than the average frame service time).

Note that for $M \geqslant 12$, although the stations become saturated, the average frame service time is only slightly

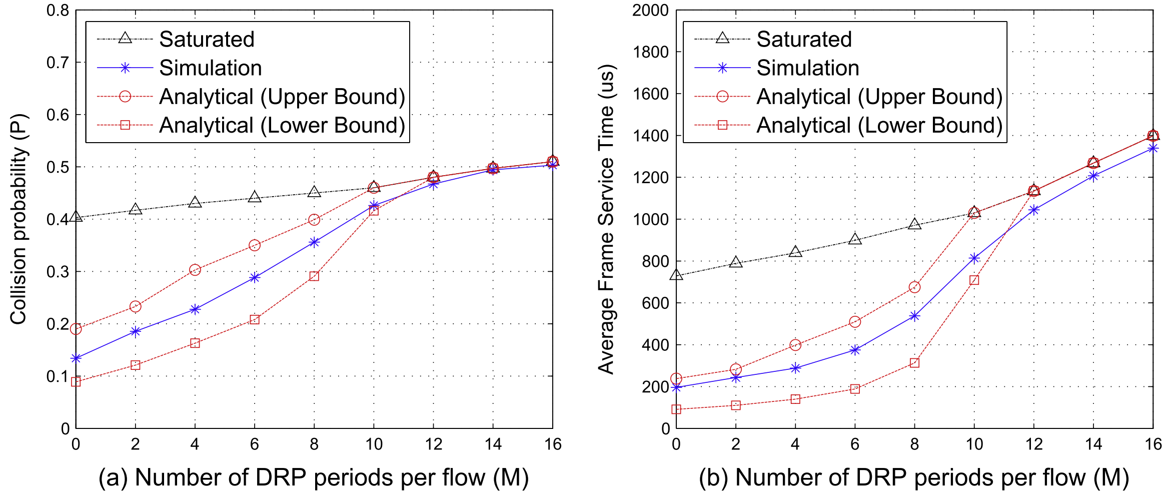


Fig. 4. Performance *vs.* M for unsaturated PCA stations.

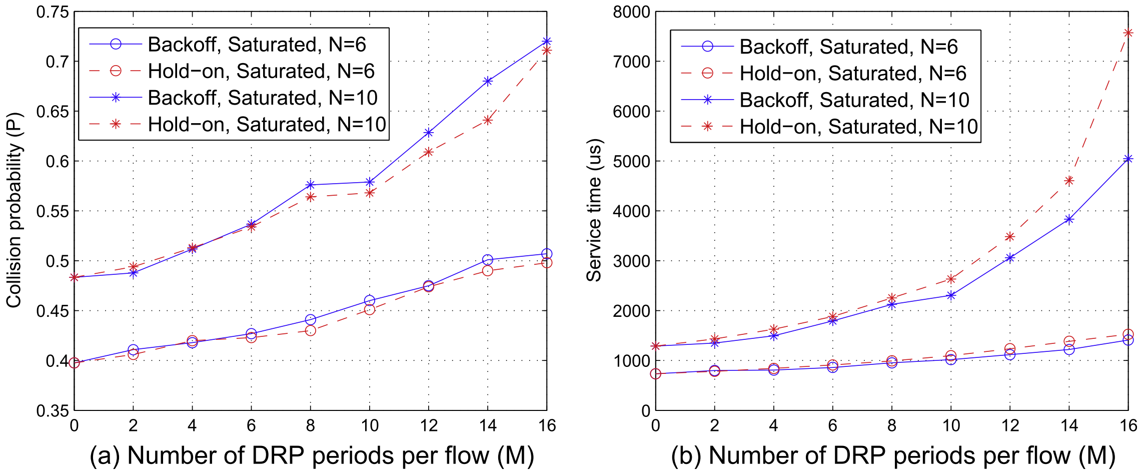


Fig. 5. Performance *vs.* M for two conflict avoidance strategies.

larger than the inter-arrival time μ, and there still exist some moments that a station is idle due to the burstiness of the traffic. Therefore, the simulation results are slightly lower than the analytical results for a real saturated station (which is always busy).

5.3. Comparison of backoff and hold-on strategies

To compare the performance of the two conflict avoidance strategies, we present the simulation results of $N = 6$ and $N = 10$ saturated stations, because the saturation throughput indicates the limit of the network capacity. The reservations of the DRP periods are the same as above.

Fig. 5a shows that the collision probability (those collisions resulting in backoff) using the backoff strategy is a little higher than that of the hold-on strategy. The difference is because of the following scenario: the tagged station obtains TXOP during the vulnerable time T_V. According to the backoff strategy, a virtual collision happens definitely and the station invokes a new stage of backoff. On the other hand, if the hold-on strategy is used, this station holds on and may successfully transmit in the slot immediately following the DRP period. In this case (*i.e.*, no other station obtains TXOP during T_V), there is no collision. This scenario is actually presented analytically by comparing (9) and (10). Obviously, P calculated in (9) (backoff strategy) is larger than that in (10) (hold-on strategy). However, given the saturated stations, the probability of this no-collision case for hold-on strategy is quite small. Consequently, in the hold-on strategy, a TXOP obtained during T_V will experience a collision after the DRP period with a high probability. Therefore, the collision probability for the hold-on strategy is slightly smaller than that of the backoff strategy.

However, the average frame service time for the backoff strategy is smaller, as shown in Fig. 5b. This is because, using the hold-on strategy, each collision results in wasted channel time of frame transaction Δ, but with the backoff strategy, a virtual collision does not waste additional channel time.

In addition, with the backoff strategy, if a station obtained TXOP during T_V, it will immediately begin the next backoff stage. So after DRP, the station may finish backoff earlier to transmit. But with the hold-on strategy, a station will delay its transmission until the slot after DRP, where it also has a high probability to have collision with other PCA stations. Then, the hold-on strategy may waste the slots in the vulnerable time for backoff.

Note that, when M increases, the percentage of the vulnerable time in the total channel time increases. Hence, the service time difference of the two strategies becomes larger. In summary, for saturated stations, the backoff strategy can lead to a higher throughput.

5.4. Simulations of video traffic

In the following, we perform simulations using real video traces to compare the performance with the two buffering architectures. The video file we used is MPEG-4 AVC encoded with the average data rate of 4.85 Mbps and the refresh rate of 30 frames/sec [17].[3] The maximum and the average video frame size are 326,905 and 20,209 bytes, respectively. The peak data rate of Group of Picture (GoP) in this file is 22.007 Mbps. There are N video streams and each stream reserves M DRP periods (one MAS per DRP period) in one superframe. Using burst transmission and block acknowledgment as shown in Fig. 1 and according to the ECMA-368 standard, an MAS can accommodate at most six video packets (with payload size of 1040 bytes including RTP, UDP, and IP headers) in our simulation. Therefore, the size of the R-Buffer in the dual-buffer architecture is six packets, while the sizes of the C-Buffer and the single-buffer are large enough to avoid packet loss due to overflow.

Because, as mentioned in Section 2.1, the compressed video traffic may be much more bursty than Poisson traffic, the collision probability and the frame service time are expected to be higher than those of Poisson traffic. Therefore,

[3] The video trace is downloaded from http://trace.eas.asu.edu/h264/mars/.

we use the upper-bound analytical model for unsaturated PCA station to estimate the performance during contention periods. Note that in the analytical model, the packets from PCA stations go through the channel only by PCA channel access. But using the proposed approach, the same flow can use both PCA and DRP to transmit packets. Thus, some packets may be transmitted in both PCA periods and DRP periods (*e.g.*, a packet is performing backoff in the PCA period. The DRP period owned by the station arrives and the packet is delivered by DRP immediately). We call the packets which have performed backoff procedure during PCA periods (no matter if it is finally delivered by PCA or DRP) *PCA packets*. The performance of these PCA packets, as shown in Fig. 6, is critical for supporting data and multimedia traffic in wireless networks.

First of all, by using the hybrid approach, with the increase of the number of MASs reserved by DRP in the superframes, *M*, more video packets are delivered by DRP and the traffic for PCA is reduced. The arrival interval (μ) for PCA packets with single-buffer is plotted in Fig. 6b. The consistent increase of μ indicates that the arrival rate of PCA packets is reduced, which is helpful to reduce the contention level during PCA periods. This is the first advantage of the hybrid approach. The analytical results plotted in the figures are calculated by plugging the new arrival interval into the upper-bound analytical model (which is presented in Section 4.2).

We examine the performance of single-buffer first. In Fig. 6a, the collision probability keeps increasing, which is because more DRP reservation periods results in more virtual collisions and also enlarge the service time of the frame. Longer service time leads to the increase of both real and virtual collisions again. Another point is that the simulation results of the video trace are larger than the analytical upper-bound. The upper-bound can work well for Poisson traffic, as shown in previous figures. But the H.264 compressed video flow is highly bursty, which results in higher collision probability.

Fig. 6b shows the average service time of PCA packets. When $M \leqslant 6$, the simulation results of video trace are larger than the analytical results for Poisson traffic. This is due to the burstiness of the video flow, similar to the case in collision probability. However, when $M \geqslant 8$, the analytical results increase rapidly, because more DRP periods increase the backoff slots significantly, as discussed above. But the service time of the video flow using hybrid approach does not increase much, which is, in particular, is much smaller than the analytical results (for a flow that can only go through PCA) when *M* is large, as indicated by the arrow 1 in the figure. The reason is as follows: by using hybrid approach, a PCA packet may be delivered by DRP. Thus, although the collision probability is high (as shown in Fig. 6a), the backoff procedure is terminated and the service time is much shorter than that if delivered by PCA. This indicates the second advantage of the hybrid approach. It is known that, when the traffic load is high, the throughput of contention-based MAC decreases dramatically due to too many collisions and long backoff time. Using the hybrid approach, appropriate reservation can help to reduce the contention among the stations, and it provides an efficient way to finish the backoff procedure earlier and increase the channel time utilization.

Now we examine the performance of dual-buffer. It can be seen that the collision probability and service time are further reduced, *e.g.*, as indicated by arrow 2 in Fig. 6b. This is expected because dual-buffer can make much higher utilization of the reservation by storing packets in the R-Buffer. So the arrival rate of PCA packets is further reduced.

It is also noticed that the collision probability and service time decrease first and then increase after $M = 6$. This is because when *M* is small, the number of packets which arrive between the station's two DRP periods is large enough to fill the R-Buffer and thus, the reservation is fully utilized and the PCA packet rate is efficiently reduced. Although more DRP periods increase the virtual collision and backoff time, the decrease on the traffic load is more significant. So the *P* and Φ can be reduced w.r.t the increase of *M*. However, when *M* is large, the R-Buffer is flushed out too frequently and there are no more packets

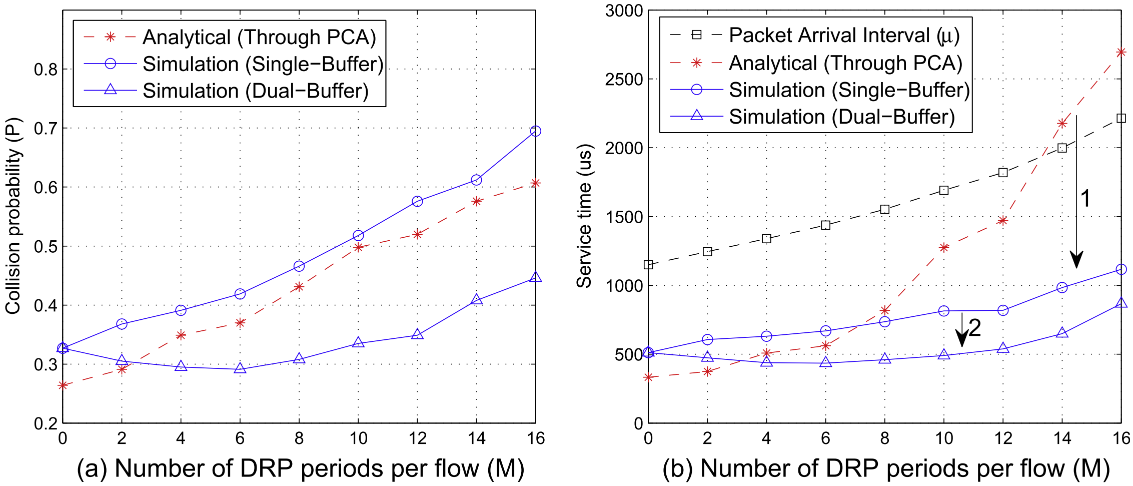


Fig. 6. Performance *vs.* *M* for video streaming.

to transmit during the DRP periods. Since the utilization of reservation time is reduced and the effect to enlarge the collision probability and service time becomes more significant, the performance degrades. In this network setting, $M = 6$ provides the optimal reservation strategy.

6. Conclusion

The main contributions of the paper are of four-fold. First, an accurate analytical model for contention access of saturated stations influenced by the reservation periods has been presented and validated. Second, for the unsaturated case, we have developed tight lower and upper bounds which converge when the stations become saturated. Third, for reservation and contention interleaved wireless systems, we have proposed two conflict avoidance strategies. Simulation results show that the backoff strategy can achieve a higher throughput when the number of reserved periods in each superframe is large. Last, for utilizing both the reservation and contention access periods, two buffering architectures have been proposed. The performance of transferring MPEG-4 video streams using the proposed hybrid approach with the two buffering architectures has been evaluated. The dual-buffer can provide a considerably better performance, thanks to the higher reservation utilization and lower contention level.

To efficiently utilize premium wireless resources and support QoS for multimedia applications, resource allocation schemes are required to be more intelligent and content-aware. As many wireless standards adopt the interleaved reservation and contention MAC protocols, the hybrid MAC approach reported in this paper is anticipated to be a key enabling technology to bridge the gap. There are many open issues beckoning for further research. For instance, how to optimize the reservation duration and pattern, considering heterogeneous data and multimedia traffic; how to make the tradeoff of hard and soft reservation to further improve the network efficiency; how to fine tune the contention access protocols at the presence of frequent reservation periods; how to flexibly utilize the hybrid MAC approach to deal with channel fading and interference, etc.

Acknowledgement

This work has been supported in part by research grants from the Natural Science and Engineering Council of Canada.

References

[1] R. Zhang, R. Ruby, J. Pan, L. Cai, X. Shen, A hybrid reservation/contention-based MAC for video streaming over wireless networks, IEEE J. Sel. Areas Commun. 28 (3) (2010) 389–398.

[2] High rate ultra wideband PHY and MAC standard, ECMA International Std. ECMA-368, December 2005. <http://www.ecma-international.org/publications/standards/Ecma-368.htm>.

[3] Q. Li, M. van Schaar, Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation, IEEE Trans. Multimedia 6 (2) (2004) 278–290.

[4] Y. Andreopoulos, N. Mastronarde, M. van der Schaar, Cross-layer optimized video streaming over wireless multihop mesh networks, IEEE J. Sel. Areas Commun. 24 (11) (2006) 2104–2115.

[5] R. Ruby, Y. Liu, J. Pan, Evaluating video streaming over UWB wireless networks, in: Proc. ACM WMUNEP'08, October 2008.

[6] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE J. Sel. Areas Commun. 18 (3) (2000) 535–547.

[7] X. Wang, G. Min, L. Guan, Performance modeling of IEEE 802.11 DCF using equilibrium point analysis, in: Proc. IEEE AINA'06, 2006.

[8] X. Ling, K.H. Liu, Y. Cheng, X. Shen, J.W. Mark, A novel performance model for distributed prioritized MAC protocols, in: Proc. IEEE Globecom'07, Washington, DC, USA, November 2007.

[9] L.X. Cai, X. Shen, L. Cai, J. Mark, Y. Xiao, Voice capacity analysis of WLAN with un-balanced traffic, IEEE Trans. Veh. Technol. 55 (3) (2006) 752–761.

[10] Y. Xiao, Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs, IEEE Trans. Wireless Commun. 4 (4) (2005) 1506–1515.

[11] C. Hu, H. Kim, J. Hou, D. Chi, S.S. Nandagopalan, Provisioning quality controlled medium access in UltraWideBand-operated WPANs, in: Proc. IEEE INFOCOM'06, Barcelona, Spain, April 2006.

[12] K.H. Liu, X. Shen, R. Zhang, L. Cai, Performance analysis of distributed reservation protocol for UWB based WPAN, IEEE Trans. Veh. Technol. 58 (2) (2009) 902–913.

[13] R. Zhang, L. Cai, Joint AMC and packet fragmentation for error-control over fading channels, IEEE Trans. Veh. Technol. 59 (6) (2010) 3070–3080.

[14] M. Daneshi, J. Pan, S. Ganti, Towards an efficient reservation algorithm for distributed reservation protocols, in: Proc. IEEE INFOCOM'10, San Diego, CA, USA, March 2010.

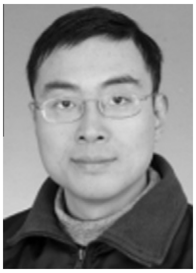
[15] Y. Tay, K. Chua, A capacity analysis for the IEEE 802.11 MAC protocol, Wireless Networks 7 (2) (2001) 159–171.

[16] Y. Lin, V. Wong, Saturation throughput of IEEE 802.11e EDCA based on mean value analysis, in: Proc. IEEE WCNC'06, April 2006.

[17] P. Seeling, M. Reisslein, Evaluating multimedia networking mechanisms using video traces, IEEE Potentials 24 (4) (2005) 21–25.

Ruonan Zhang is currently an associate professor of Schools of Electronics and Information at the Northwestern Polytechnical University, Xi'an, Shaanxi, China. He received the B.S. (2000) and M.S. (2003) degrees from Xi'an Jiaotong University (China) and Ph.D. degree (2010) from University of Victoria (Canada), respectively. He was with Motorola Inc. and later with Freescale Semiconductor Inc. in Tianjin, China, from 2003 to 2006, working on IC architecture and application design. His current research interests include cross-layer design and optimization for wireless networks.

Lin Cai 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 and then an Associate 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 has been awarded the NSERC Discovery Accelerator Supplement Grant in 2010. 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).



Jianping Pan is currently an assistant professor of computer science at the University of Victoria, British Columbia, Canada. He received his Bachelor's and Ph.D. degrees in computer science from Southeast University, Nanjing, China, and he did his postdoctoral research at the University of Waterloo, Ontario, Canada. He also worked at Fujitsu Labs and NTT Labs. His area of specialization is computer networks and distributed systems, and his current research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security. He is a senior member of the ACM and a senior member of the IEEE.

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 University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on resource management in interconnected wireless/wired networks, UWB wireless communications networks, wireless network security, wireless body area networks and vehicular ad hoc and sensor networks. He is a co-author of three books, and has published more than 400 papers and book chapters in wireless communications and networks, control and filtering. He served as the Technical Program Committee Chair for IEEE VTC'IO, the Tutorial Chair for IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the 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; Editor-in-Chief for Peer-to-Peer Networking and Application; Associate Editor for IEEE Transactions on Vehicular Technology; Computer Networks; and ACM/Wireless Networks, etc. He has also served as Guest Editor for IEEE JSAC, IEEE Wireless Communications, IEEE Communications Magazine, and ACM Mobile Networks and Applications, etc. He received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 and 2008 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, and a Distinguished Lecturer of IEEE Communications Society.

Please cite this article in press as: R. Zhang et al., Resource management for video streaming in ad hoc networks, Ad Hoc Netw. (2010), doi:10.1016/j.adhoc.2010.08.012