Investigation of the Block ACK Scheme in Wireless Ad-hoc Networks

Tianji Li[1,*,†], Qiang Ni[2,‡], and Yang Xiao[3§]
[1]Hamilton Institute, National University of Ireland at Maynooth, Co. Kildare, Ireland.
[2]Electronic & Computer Engineering, Brunel University, UB8 3PH, West London, UK.
[3]Computer Science Department, University of Memphis, Memphis, TN 38152, U.S.A.

Summary

A Block Transmission and Acknowledgement (BTA) scheme, also called Block ACK, has been proposed in the IEEE 802.11e wireless LAN (WLAN) specification to improve efficiency of the medium access control layer. The idea of the BTA scheme is to transmit multiple data frames followed by only one acknowledgment frame in a transmission block. In this paper, we present a theoretical model to evaluate the saturation throughput for the BTA scheme under error channel conditions in the ad-hoc mode, validated with simulations. We show some advantages of BTA over the legacy MAC, and analyze how to select a proper number of frames for each transmission block. Results show that BTA is particularly effective in very high-speed wireless networks, and it is important that the number of frames in each block is negotiated before transmissions to provide better efficiency.

KEYWORDS: medium access control (MAC); wireless ad-hoc networks; Block ACK (BTA)

1 Introduction

Wireless ad-hoc networks have received significant attentions, partially due to their flexibility and low cost. Most of researches in wireless ad-hoc networks focus on the layer 3, i.e., routing protocols. Whereas, medium access control (MAC) is another major aspect for designing wireless ad hoc networks. Currently, the most popular MAC protocol for ad-hoc is the one designed for the IEEE 802.11 Wireless Local Area Networks (WLAN) [1, 2], mostly due to the fact that it is the only available protocol in reality although there are plenty protocols proposed in the literature for academic research purposes, and they are not likely implemented in the real world.

The IEEE 802.11 standard [3] defines two MAC access methods: a distributed coordination function (DCF) and an optional point coordination function (PCF). This paper focuses on using the DCF in ad hoc mode, while the PCF is hardly implemented in reality.

The IEEE 802.11 DCF protocol adopts carrier sense multiple access with collision avoidance (CSMA/CA) with binary exponential backoff, in which overhead of the MAC and Physical (PHY) layers is a major obstacle to system efficiency. The overhead includes backoff time, inter-frame spaces , acknowledgement frame (ACK), and MAC and PHY layer headers, etc [4, 5].

Much work in IEEE 802.11 has been done to optimize the backoff process [6, 7, 8, 9]. However, the backoff process cannot avoid possible collisions and idle slots due to its randomized characteristic. On the other hand, even without the problem caused by the randomized backoff, the overhead is not negligible, and it is particularly important when data rates are very high [4, 5].

To mitigate the impact of the overhead, a Burst ACK scheme has been proposed in [10]. In the Burst ACK scheme, only the first frame in a burst contends for the channel access. Once a station obtains a transmission opportunity, it sends out multiple frames in a burst with-

*Correspondence to: Tianji Li, Hamilton Institute, National University of Ireland at Maynooth, Co. Kildare, Ireland, Tel: +353-1-7086273, Fax: +353-1-7086269. Email: Tianji.Li@nuim.ie

†The work of Tianji Li was supported by the Science Foundation Ireland under Grant 03/IN3/I396.

‡Qiang Ni is now with Electronic & Computer Engineering, School of Engineering and Design, Brunel University, West London, UK. Part of his work was done while he was with the Hamilton Institute.

§Yang Xiao is now with Computer Science Department, The University of Memphis, Memphis, TN 38152 USA, E-mail: yangxiao@ieee.org.

out contending the channel again. Each frame is immediately acknowledged by an ACK frame. Thus, there are multiple data and multiple ACK frames in a transmission burst.

Recently, a new scheme based on the Burst ACK is proposed in 802.11e, which is known as block ACK (BTA) [11]. In the BTA scheme, multiple data frames are sent out when a channel access chance is obtained, and they are acknowledged by only one ACK frame at the end of the transmission block. In this way, higher efficiency is expected. Other very related work includes concatenation schemes [12, 13], packing and aggregation schemes [14, 15] and the AFR scheme [16, 17].

Although many analytical models have been proposed for the legacy DCF scheme (e.g., [18], [19], [20]), few prior work has studied the performance for this BTA scheme. In [5], authors investigate the ideal case throughput for the BTA scheme. The saturation throughput of the BTA scheme in an infrastructure network is studied in [21] with the assumption that the channel is error-free.

To the best of our knowledge, none of the existing work has focused on the ad-hoc performance of the BTA scheme in a noisy environment. Thus, we propose an analytical model called BTA-MODEL which is an extension of the DCF model proposed in [18]. The key observation that enables our extension is that each transmission block in BTA can be treated as a single frame of the DCF. The validation of this BTA-MODEL is verified through *NS-2* simulations. Using this model, we first show advantages of BTA over the legacy MAC, then analyze how to select a proper block size for the BTA scheme, and describe a protective mechanism in which the number of frames in each block is negotiated before transmissions.

The rest of this paper is organized as follows. The legacy DCF and Burst ACK schemes are introduced in Section 2. In Section 3, we introduce the BTA scheme. The analytical model for BTA is then described in Section 4. Section 5 presents our implementation of the BTA scheme and introduces the corresponding analysis based on the results from both simulations and the proposed model. Finally, Section 6 concludes this paper.

2 The Legacy Schemes

2.1 The DCF Scheme

In the legacy DCF scheme, a station (STA) can transmit a frame after observing an idle medium for a distributed inter-frame space (DIFS) plus a backoff duration. If this frame is received correctly, then the destination within the same range sends back an ACK frame after a short inter-frame space (SIFS) period, which is the interval needed by the physical (PHY) layer to turn from the receiving state to the transmission state. All the other STAs defer the channel contention until the end of the ACK transmission. After that, the destination and all the other STAs defer a DIFS duration before counting down their backoff counters for the next round of transmission.

Possible collisions and transmissions errors make the MAC layer protocol complicated. In this paper, we define a *collision* as the event that at least two STAs start transmission at the same time and the receivers can not decode frames correctly. We define an *error* as the event satisfying the following two conditions at the same time. First, there is one and only one STA transmitting but the channel is so noisy that the destination can not decode the whole frame successfully; second, although the PHY layer has detected errors, it still completes the reception and transfers the received error frame to MAC, which detects the error by using checksum. According to this definition, an *error* in this paper is a MAC layer frame transmission error instead of a normally used PHY concept[1].

In the case of collisions or errors, all the STAs except the sender defer their own transmission attempts for an EIFS duration. The duration of EIFS is the sum of a SIFS, a DIFS and an ACK transmission interval, i.e., $T_{EIFS} = T_{SIFS} + T_{PHYhdr} + T_{ACK} + T_{DIFS}$. The sender waits for the potential ACK until an ACK timeout event, and then defers a backoff interval before a retransmission. Notations used in this paper are listed in Table 1.

The total length of the backoff period is the product of the slot time [2] and a random number uniformly

[1] In reality, errors may be also due to collisions if the PHY layer is able to receive the transmission from multi-users simultaneously or there are hidden terminals. Then an *error* can be defined as the event that although the receiver's PHY completes a reception, the frame that MAC receives still contains errors. A *collision* can be defined as the event that the receiver can detect the coming signals but the reception is always interrupted.

[2] Slot time is PHY dependent. The length of the slot time is 9 μs

n	Number of STAs
$T_{\overline{CW}}$	Average backoff duration
T_{SIFS}	Time duration of SIFS
T_{DIFS}	Time duration of DIFS
T_{EIFS}	Time duration of EIFS
T_f	Time duration to transmit a frame in BTA
T_{data}	Time duration to transmit a frame in DCF
T_{bar}	Time duration to transmit a BAR frame
T_{ba}	Time duration to transmit a BA frame
T_{ack}	Time duration to transmit an ACK frame
T_{PHYhdr}	Time duration for PHY header
δ	Propagation delay
σ	Idle slot duration
L_{pld}	MAC layer payload size in BTA (bytes)
L_f	MAC layer frame size in BTA (bytes)
L_{CRC}	CRC size (bytes)
L_{data}	MAC layer frame size in DCF (bytes)
L_{ack}	MAC layer ACK frame size (bytes)

Table 1: Notations

chosen from the range of $[0, CW-1]$, where CW is the current contention window size when the back-off number is generated. Note that the backoff period for one station may overlap several transmission blocks. CW is doubled after each failed transmission until the maximum contention window size CW_{max} is reached. After each successful transmission, CW is reset to the minimum contention window size CW_{min}, where $CW_{min} \leq CW \leq CW_{max}$.

In a CSMA/CA-based scheme, the overhead of MAC and PHY is the main reason for system inefficiency. To show the inefficiency caused by overhead, we first calculate and present the MAC efficiency of the legacy DCF based on an ideal case assumption. In the ideal case, the channel is assumed perfect, i.e., neither errors nor collisions occur, and in any transmission cycles, there is only one active STA which always has backlogged[3] frames to transmit. The receiver responds with ACKs, and the other STAs only sense the channel and wait. We can define the average length of the backoff as $T_{\overline{CW}} = (CW_{min}-1)\cdot\sigma/2$, where σ stands for the idle slot duration. T Then, the ideal throughput S_{ideal}^{DCF} can be defined as follows [5]:

$$S_{ideal}^{DCF} = \frac{8 \cdot L_{data}}{T_{DIFS} + T_{\overline{CW}} + T_{data} + T_{SIFS} + T_{ACK} + 2\delta}. \quad (1)$$

for IEEE 802.11a [25]. In this paper, we use a generic slot time, which is the same as that defined in IEEE 802.11 if the channel is idle, but if the channel is busy, our generic slot time is defined as the duration in which the channel is sensed busy.

[3]A frame is said to be backlogged if it is in the queue between the MAC and its upper layer waiting to be transmitted.

SIFS (μs)	16
Slot time (σ) (μs)	9
DIFS (μs)	34
PHYhdr (μs)	20
CW_{min}	16
CRC (bits)	32
Propagation delay (δ) (μs)	1
Symbol delay (μs)	4
PHY rate (Mbps)	54·k (k=1,2,3,...)
Retry limit	4
Frame size (bytes)	1024

Table 2: The MAC/PHY parameters used in this paper.

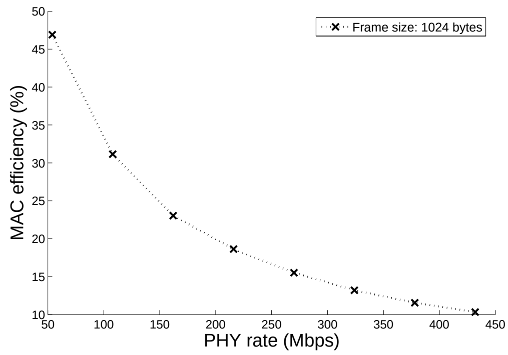


Figure 1: The MAC efficiency of the legacy DCF in the ideal case with a 1024-byte frame size. The x-axis represents the PHY rate. The y-axis represents the ratio of the ideal throughput to the PHY rate.

The parameters are listed in Table 2.

Using Equation (1), we illustrate in Fig. 1 the MAC efficiency while the PHY rate is increased from 54 Mbps to 432 Mbps. Here, the MAC efficiency represents the ideal throughput normalized to the PHY rate. As illustrated in the figure, the efficiency decreases dramatically as the PHY rate increases. Moreover, even though the PHY rate is infinitely high, the MAC throughput is still bounded [5].

2.2 The Burst ACK Scheme

In the Burst ACK scheme proposed in [10], only the first frame in a burst contends for the channel access. Once the STA obtains the transmission opportunity, it sends out multiple frames in a burst without contending the channel again. As illustrated in Fig. 2(b), each frame is immediately acknowledged by an ACK frame.

Thus, there are multiple data frames and multiple ACK frames in a transmission burst. It may be more effective than DCF due to the fact that the frames and ACKs share a same transmission opportunity, which decreases the overall probability of collisions.

3 The 802.11e Block ACK Scheme

A Block ACK (BTA) scheme is proposed in the IEEE 802.11e specification [11]. In the BTA scheme, a block of frames sent to the same destination is allowed to be transmitted without being acknowledged, and each frame is back-to-back separated by a SIFS period. Thus the backoff process is generated for a transmission block instead of a single frame, shown in Fig. 2(c). After the block, a block acknowledge request (BAR) frame is initiated by the sender to enquire which frames have been received successfully, and then a block acknowledge (BA) frame is sent back by the receiver to answer this enquiry.

Upon receiving the BA frame correctly, the sender should defer a DIFS interval and a backoff process before sensing the channel again. Meanwhile, all the other STAs should wait until the end of the BA transmission, and then defer another DIFS interval before counting down their backoff counters for the next round of transmission.

If two or more STAs start transmissions in a same slot, a collision occurs. Each of them sends out a whole block and a BAR frame, and then waits for the BA frame. The receivers shall not send back the BA frames if they can detect the collisions; otherwise more than one BA frames will be sent back to the senders. In neither cases, the senders can receive the BA frames successfully, and thus the senders have to retry their transmissions.

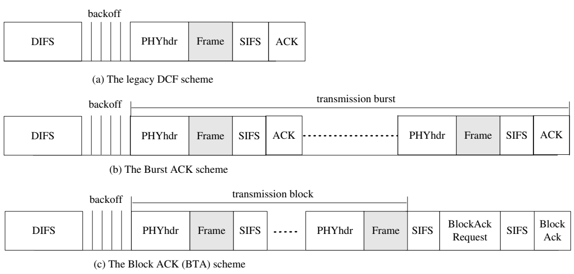


Figure 2: The three schemes considered in this paper.

In the erroneous case, a sender sends out a whole block and a BAR frame. The receiver then sends back a BA frame to indicate which frames are successfully. If the sender receives the BA frame successfully, those correctly transmitted frames in the block will be removed from the sending queue and a new block will be constructed for the next round of transmission.

3.1 Frame Formats

Fig. 3(a) shows the format of a BAR frame. There are two new fields in the BAR frame. The *BAR control* field is shown in Fig. 3(b). This field is used for quality-of-service negotiation between MAC and its upper layer. The *Block ACK Starting Sequence Control* field is shown in Fig. 3(c). The last 12 bits of this field are used to record the first frame's sequence number in a block, the first 4 bits are reserved for further usage.

2	2	6	6	2	2	4
Frame control	Duration	Receiver Address	Sender Address	BAR Control	Block ACK Starting Sequence Control	CRC

(a) Block ACK Request packet

12 bits	4 bits
Reserved	TID

(b) BAR Control field

4 bits	12 bits
Fragment Number	Starting Sequence Number

(c) Block ACK Starting Sequence Number field

Figure 3: Format of the Block ACK Request frame.

To inform the sender which frames have been lost in a block, a *Block ACK Bitmap* field is designed in the BA frame as illustrated in Fig. 4. It is a 128-byte field, and thus it can support up to $128 \times 8 = 1024$ frames in a single block. The *Block ACK Starting Sequence Control* field is used to indicate to which BAR this BA frame responds.

2	2	6	6	2	2	128	4
Frame control	Duration	Receiver Address	Sender Address	BA Control	Block ACK Starting Sequence Control	Block ACK Bitmap	CRC

Figure 4: Format of the Block ACK frame.

3.2 Discussions

We have the following observations and discussions.

- BTA and previous schemes differ in the following ways. Firstly, the unit of transmission in the BTA scheme is a block, which is consisted of multiple data frames and one ACK. The unit of transmission in the DCF is a data frame and an ACK frame, and in the Burst ACK scheme is a burst which contains multiple data frames and multiple ACKs. Therefore, the BTA scheme is expected to be more effective. Secondly, in previous schemes, a data frame is acknowledged immediately by an ACK. In BTA, however, a modified sending queue and a receiving queue are required to accommodate block transmissions.

- In the case of a collision, a whole block will be retransmitted. Therefore, a protective mechanism is needed to solve this problem. In such a mechanism, the number of frames (the block size) in each block is negotiated between the sender and the receiver. In an infrastructure mode, the protection can be accomplished by the access point (AP). AP periodically broadcasts the start time and the block size to all the STAs. In an ad-hoc network, however, the protection has to be done in a distributed manner. To this aim, IEEE 802.11e [11] proposes two ways as follows. First, a similar method as Request-To-Send (RTS)/Clear-To-Send(CTS) can be used, i.e., before each block transmission, the sender sends an *Add Block ACK Request* frame to the receiver which should respond with an ACK, and then the receiver sends an *Add Block ACK Response* frame to the sender. Another and better solution is to acknowledge each block's first frame, in which the block size is carried. Interestingly, the protective mechanisms proposed for the block size can also be used to mitigate the collision problem mentioned before.

- BTA can be used as a solution for the multi-rate fairness problem in CSMA/CA-based networks. Recently, [23] has shown that a CSMA/CA-based network distributes transmission probabilities fairly amongst all the STAs. In networks where STAs have different PHY rates, this characteristic is actually not fair for faster STAs in the sense that they should be able to achieve higher throughput than the slower ones. BTA can be used to mitigate this problem by transmitting more frames back-to-back for faster STAs than for slower ones.

- Finally, it can be seen that the BTA scheme operates in a similar way to the legacy DCF. In particular, we may treat a block in the BTA scheme as a frame in the DCF because both of them are considered as a unit of operation. This understanding suggests that it is possible to extend previous analysis which was designed for the legacy DCF to study the BTA scheme. The similar technique has also been used in [13].

4 An Analytical Model for BTA Scheme

In this section, we present an analytical model to compute the saturation throughput for the BTA scheme under an error channel.

We consider an ad-hoc network where all the STAs can hear each other, i.e., one-hop ad hoc network. In such an area, collisions occur only when at least two STAs start transmissions at the same time. A transmission error occurs when there is only one STA which is transmitting in a given slot, but the transmission can not be received correctly because of channel noise. We assume that the PHY headers are always transmitted successfully given the fact that they are usually transmitted at the basic hence the safest rate [3]. We also assume that the transmissions of the BAR and BA frames are always successful.

4.1 Saturation Throughput

Based on previous work [18], [19] and [20], we have designed an analytical model for the BTA saturation throughput S_{BTA}, which is defined as the payload size of the successfully transmitted frame $E[L_{pld}]$ in an expected slot duration $E[T]$.

$$S_{BTA} = \frac{E[L_{pld}]}{E[T]}. \tag{2}$$

We first compute the expected slot duration $E[T]$. There are four types of durations in the BTA scheme as shown in Fig. 5.

- If none of the STAs transmit any frames, they all wait for a duration $T_i = \sigma$, where σ corresponds to the idle slot interval.

- Let T_S denote the duration during which a whole block is transmitted successfully. In this case, only one STA transmits frames and its transmission is always successful. The channel state shall be kept busy in a duration which is equal to the duration of a block of frames' transmission plus $(N_b - 1)$ SIFSs, a BAR and a BA transmission, where N_b denotes the block size.

- Let T_E be the duration in which at least one frame in a block is corrupted due to the channel errors. The sender shall not stop the transmission and the receiver shall respond with a BA frame. The other STAs defer a block and a DIFS duration.
- Let T_C denote the collision duration in which at least two STAs start transmission simultaneously. In this case, no BA frames are initiated by the receivers. All the other STAs except the senders and the receivers defer for an EIFS ($T_{EIFS} = T_{SIFS} + T_{PHYhdr} + T_{ba} + T_{DIFS}$) interval.

The slot durations can be expressed as follows:

$$\begin{aligned}
T_I &= \sigma \\
T_S &= N_b \cdot (T_f + T_{SIFS}) + T_{DIFS} + \\
&\quad (T_{bar} + T_{SIFS} + T_{ba}) + (N_b + 2)(T_{PHYhdr} + \delta) \\
T_E &= T_S \\
T_C &= N_b \cdot (T_f + T_{SIFS}) + T_{EIFS} + \\
&\quad (T_{bar} + T_{SIFS} + T_{ba}) + (N_b + 1)(T_{PHYhdr} + \delta).
\end{aligned}$$

We then turn to calculate the corresponding possibilities for the slot durations. Let τ and n denote a STA's transmission probability in a slot and the number of STAs in the system, respectively.

First, for an idle slot, a single STA does not attempt transmission with probability $(1 - \tau)$, and then all the n STAs in the system keep silent with probability $P_I = (1 - \tau)^n$ as shown in Equation (3).

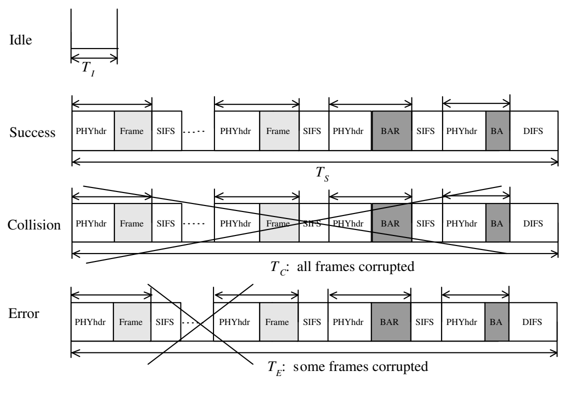


Figure 5: Time durations in the BTA scheme

Second, let p_e^{bta} denote a single STA's error probability for an entire block, and then the successful probability can be expressed as in Equation (4). Similarly, we get the system error probability P_E in Equation (5).

Finally, since these four events (idle, success, collision and error) are mutually exclusive [27], collision probability for a system can be defined as in Equation (6).

$$P_I = (1 - \tau)^n \tag{3}$$
$$P_S = n \cdot (\tau(1 - \tau)^{n-1}) \cdot (1 - p_e^{bta}) \tag{4}$$
$$P_E = n \cdot (\tau(1 - \tau)^{n-1}) \cdot p_e^{bta} \tag{5}$$
$$P_C = 1 - P_I - P_S - P_E. \tag{6}$$

Let p_e denote the frame error rate (FER) of a frame. The probability p_e^{bta} can be expressed as:

$$p_e^{bta} = 1 - (1 - p_e)^{N_b}. \tag{7}$$

p_e can be computed if the bit error distribution is given. We use the discrete-time, memory-less Gaussian channel as an example. In such a channel, the bit errors independently and identically distribute over a frame [22]. Let L_f and p_b denote the frame size and the bit error rate (BER), respectively. p_e is defined as:

$$p_e = 1 - (1 - p_b)^{L_f}, \tag{8}$$

where the p_b is assumed to be known by the MAC layer. In reality, it can be measured by the PHY layer. If the p_b measurement is not available, p_e can be measured instead since it is easier for the PHY layer.

Although the memory-less Gaussian model is unable to capture the fading characteristics of the wireless channel, it is widely used in modelling wireless MAC layers since the focus here is the MAC protocol itself. Moreover, if interleaving is employed, the BER will become Gaussian-like.

So far we have known all the variables except probability τ in Equations (3-6). Let p_f denote the probability of doubling contention window after a failed transmission. The probability τ can be expressed as a function of p_f, and we can find another function of τ for p_f. Both of them are obtained from a Markov chain that is similar to the one in Bianchi's paper [18]. We will explain this Markov chain in Appendix.

Finally, all the variables in Equations (3-6) have been defined. The saturation throughput S_{BTA} can then be expressed as:

$$S_{BTA} = \frac{P_S \cdot N_b \cdot L_f + P_E \cdot E[L]}{P_I T_I + P_S T_S + P_E T_E + P_C T_C}, \tag{9}$$

where $E[L]$ stands for the expected frame size successfully transmitted in an erroneous case. Let i denote the number of the corrupted frames. Based on the same time-less Gaussian assumption, $E[L]$ can then be expressed as:

$$E[L] = \sum_{i=1}^{N_b} \binom{N_b}{i} \cdot (p_e)^i \cdot (1-p_e)^{N_b-i} \cdot (N_b - i) \cdot L_f. \quad (10)$$

5 Evaluation

We implemented the BTA scheme in the network simulator *NS-2* [24] to validate our analytical model. The simulation parameters are listed in Tables 2 and 3.

First, in the *NS-2* version 2.27, the PHY headers are transmitted with the same rate as the data frames. However, the IEEE 802.11a [25] specifies that the PHY headers should be transmitted with a low data rate but within $20\mu s$ no matter what the data part length is. We revised the *NS-2* codes according to the IEEE 802.11a specification. Second, all the STAs are placed within the same range so that there are no hidden terminals. Furthermore, we need to ensure that all the STAs achieve the same throughput because all of them are modeled by a single Markov chain in the BTA-MODEL Note that a same throughput for all the STAs is also required in Bianchi's model [18]. To gauge whether this fairness goal is reached in the *NS-2* simulations, we use the fairness index I, a real value between 0 and 1, defined as follows [26]:

$$I = \frac{(\sum_{i=1}^{n} S_i)^2}{n \cdot \sum_{i=1}^{n} S_i^2}, \quad (11)$$

where n stands for the number of STAs and S_i denotes the throughput of STA i. When each STA achieves exactly the same throughput, I is equal to 1. In our simulations, we run each test for a duration that is long enough to obtain a fairness index I close to 1. If only one STA happens to dominate the channel entirely, I approaches $1/n$.

Finally, we introduce our implementation in the following. A *bitmap array*, a sending queue (*Sq*) and a receiving queue (*Rq*) are used. The *bitmap array* is for recording the number of frames that have been transmitted successfully. The *Sq* and the *Rq* are used to save frames temporarily at the MAC layer. For convenience, let h_{Sq}, t_{Sq}, h_{Rq}, and t_{Rq} denote the head of the *Sq*, the tail of the *Sq*, the head of the *Rq*, and the tail of the *Rq*, respectively.

The sender stores a frame from the upper layer at the t_{Sq}, and checks whether N_b (the block size) frame have been transmitted. If so, it constructs a BAR frame at the MAC layer and transmits it. Otherwise, the first frame at the h_{Sq} shall be popped out and be transmitted.

On reception of a data frame f_j, the receiver checks its correctness and updates accordingly the *bitmap array* whose length is equal to N_b. Then f_j is appended at the t_{Rq} if it has not been received before. If f_j has been in the *Rq* but marked as 'corrupted', the receiver updates its flag.

Upon receiving a BAR frame, the receiver responds with a BA frame containing the *bitmap array*. Then the *bitmap array* should be reset for the next round of receiving, and all the correctly received frames in the *Rq* are transferred to the upper layer.

After receiving a BA frame, the sender removes all the frames that have been received successfully from the *Sq*. The CW size will be reset for both successful and erroneous transmissions.

In the case of collisions, receivers do not initiate the BA frames. Then after a transmission block, a sender waits until the BAR timeout and retransmits the entire block.

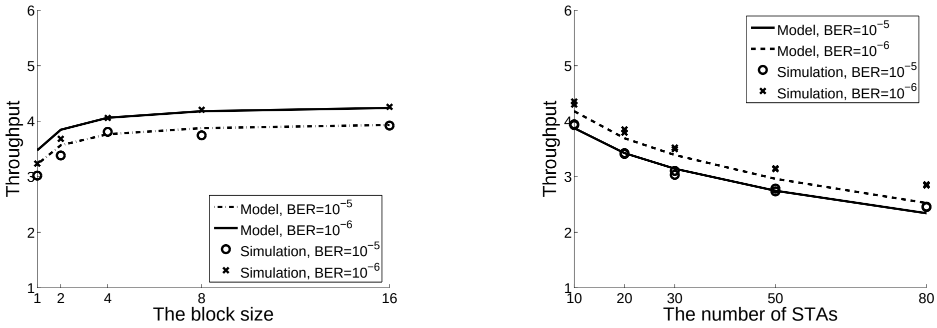
5.1 Results

In this section, we introduce the results from the *NS-2* simulations and the theoretical analysis. First, we compare the results from the analysis and the simulations to show that the model we developed in Section 4 is correct. Second, we compare the BTA scheme with the legacy DCF scheme to show the superiority of the former one. Third, using the theoretical model, we analyze the most important characteristic of the BTA scheme, i.e., the block size of BTA.

5.1.1 Model Validation

We validate the BTA model in two ways. Firstly, a 10-STA ad-hoc WLAN is simulated, in which each STA has a PHY rate of 6Mbps and a UDP traffic rate of 6Mbps. BAR and BA frames are also transmitted at 6Mbps. Other parameters are listed in Tables 2 and 3. We plot the results from the simulation and the model in which the block size is increased from 1 to 16. The results are illustrated in Fig. 6(a), which shows that that the results of our model matches well the results of the simulations.

Secondly, we fix the block size and increase the number of STAs in this network. The corresponding results are plotted in Fig. 6(b). Again, our proposed model works well as the channel becomes highly loaded.

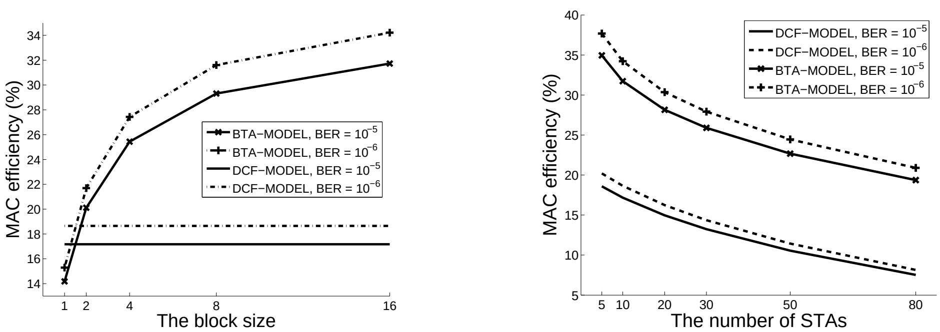


(a) (b)

Figure 6: Model validation: The parameters are listed in Table 2 and 3.

	Fig. 6(a)	Fig. 6(b)	Fig. 7(a)	Fig. 7(b)	Fig. 8(a)	Fig. 8(b)
Number of STAs	10	varied	10	varied	10	10
Block size	varied	8	varied	16	varied	varied
Application rate (Mbps)	6	6	-	-	varied	54
PHY data rate (Mbps)	6	6	216	216	varied	54
BAR/BA rates (Mbps)	6	6	216	216	varied	6
BTA Sq (frames)	20	20	-	-	-	-
BTA IFQ (frames)	10	20	-	-	-	-

Table 3: The parameters used in the simulations and the analysis.



(a) (b)

Figure 7: (a) Throughput: BTA-MODEL vs DCF-MODEL while increasing the block size. (b) Throughput: BTA-MODEL vs DCF-MODEL while increasing the number of STAs. The y-axis represents the ratio between the MAC throughput and the PHY rate. The MAC/PHY parameters are listed in Tables 2 and 3.

5.1.2 Comparison with DCF

As a first application of the designed model, we use it to compare BTA with the legacy DCF scheme. To this end, a model for the legacy DCF scheme is required. We use the DCF-MODEL that has been developed and validated in our previous work [20].

In both schemes, the definitions of collision and error are the same, and only the data frames can be corrupted in the case of errors. The BAR and BA frames in the BTA scheme and the ACK frames in the legacy DCF are always transmitted correctly.

Meanwhile, the DCF-MODEL has two differences to the BTA-MODEL. First, ACK duration is used instead of the BAR and BA durations. Second, EIFS rather than DIFS is deferred for the erroneous transmission, and $T_{EIFS} = T_{SIFS} + T_{PHYhdr} + T_{ack} + T_{DIFS}$. Thus, the slot durations for the DCF-MODEL are[4]:

$$\begin{aligned} T_I &= \sigma \\ T_S &= T_{data} + T_{SIFS} + T_{ack} + T_{DIFS} + 2(T_{PHYhdr} + \delta) \\ T_E &= T_{PHYhdr} + T_{data} + T_{EIFS} + \delta \\ T_C &= T_E. \end{aligned} \quad (12)$$

The corresponding probabilities are listed in (13). Then the saturation throughput for the legacy DCF (S_{DCF}) can be expressed as in Equation (14). Readers please read [20] for details of the DCF-MODEL under an error channel.

$$\begin{aligned} &P_I = (1-\tau)^n \\ &P_S = n \cdot (\tau(1-\tau)^{n-1}) \cdot (1-p_e) \\ &P_E = n \cdot (\tau(1-\tau)^{n-1}) \cdot p_e \\ &P_C = 1 - P_I - P_S - P_E. \end{aligned} \quad (13)$$

$$S_{DCF} = \frac{P_S \cdot L_{data}}{P_I T_I + P_S T_S + P_E T_E + P_C T_C}. \quad (14)$$

To compare both schemes, we first use a 216Mbps PHY rate for both of them. In Fig. 7(a), as the block size increases, BTA achieves considerable higher throughput than that of the DCF. If there is only one frame in a block, the performance of BTA is lower than DCF. The reason is that BTA has two control frames (BAR and BA) for only one frame, but DCF uses only one (ACK) for this purpose.

Then, we compare them in ad-hoc networks while increasing the traffic load. In this simulation, each STA i has a single UDP link to STA $i+1$. The results are illustrated in Fig. 7(b). It can be seen that the BTA scheme always achieves higher throughput than the legacy DCF as the number of STAs increases from 5 to 80.

[4]The DCF-MODEL in [20] has five time durations because transmission errors of ACK frames are also considered.

5.1.3 The Block Size

An important question for BTA is how to choose a proper block size, or whether it is possible to find an *optimal block size*, which always provides the best performance. We investigate this issue in this section.

Firstly, we plot the results in Fig. 8(a) showing the performance improvement of BTA over DCF under saturated channel conditions, with an increased block size from 1 to 64 bytes. Two interesting observations are obtained: First, a block size of 16 is good enough, where larger block sizes (e.g., 32 and 64) introduce minor improvements. Thus, we recommend that the optimal block size in a saturated network is 16 under the current parameter setting. Second, for STAs with 6Mbps, 54Mbps and 108Mbps, the BTA scheme is not effective because it only achieves negligible improvement (less than 10%) over the legacy DCF scheme. Note that the frame size used in this example is 1024 bytes. With smaller frames, the improvement of BTA over DCF is higher. This second observation is due to the fact that one advantage of BTA comes from reducing the overhead caused by multiple ACKs, whose impacts become comparatively larger when the frame size is smaller, or when the PHY data rate is very high. Thus, the conclusion that can be drawn here is that BTA is very effective in a high-speed network.

Secondly, we turn to exploit the optimal block size in a non-saturated case where frames are generated infrequently. As in Fig. 8(b), the x-axis represents the block size, the three curves represent the case where there are 4, 8, and 16 frames in average available in the Sq before a block transmission, respectively. Each curve has a peak which represents there are n frames in each block while the block size is also n. The left side on the peak represents the results when there are more frames in the sending queue than the block size, the leftist parts of the three curves are overlapped since they have the same value; the right side on the peak represents the cases where there are less frames in the sending queue than the expected block size.

Note that, these curves decrease rapidly on the right side, which means if we use a larger block size than the available number of frames in the Sq, the efficiency of BTA decreases dramatically. To further explain this

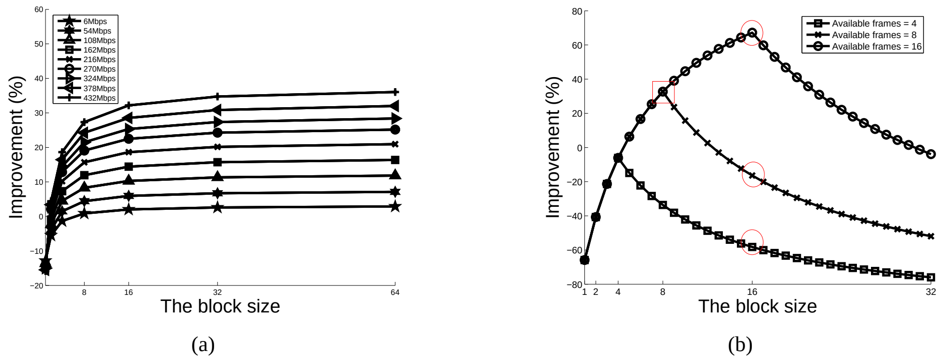


Figure 8: (a) Optimal number of frames in a block for a saturated case. (b) Optimal number of frames in a block for a non-saturated case. The parameters are listed in Tables 2 and 3.

phenomenon, let us look at the three points in circles. All these three points are for the cases where the block size is 16, but the upper point represents the improvement of BTA over DCF when there are 16 frames available in the Sq, the middle and the lowest ones for 8 and 4 frames available, respectively. It can be seen that for the latter two cases, the block size of 16 leads to negative improvement (-20% and -60%). That is, the BTA scheme should not be used. In another word, a block size of 16 is not the optimal value anymore in the non-saturated cases. Similarly, 4 or 8 are not qualified for the optimum neither if there are not enough frames available.

However, this figure has already told us the way to optimize the efficiency, that is to find the peaks for each curves. For example, if there are always 8 frames available, due to the sharp peak of the curve, the only promising block size is 8 (the point in the square), thus a mechanism should be designed to negotiate the block size between the sender and the receiver, and this should be done by a per-block basis.

Two methods has been proposed in 802.11e for this purpose [11]. First, before each block transmission, a RTS/CTS-like mechanism is used to exchange the block size. Second, the first frame in each block must be acknowledged by a special ACK frame. In this second method, the first frame needs to carry the block size information, and the special ACK must be received before transmitting the other frames in the block.

5.1.4 Summary

First, the optimal block size is recommended to be 16 bytes if channel is saturated, i.e., there are always enough frames in a sender queue.

Second, in situations where there is no guarantee that enough frames are available, a protective mechanism is required for negotiating the actual block sizes. This is of vital importance for BTA in ad-hoc networks.

6 Conclusion

In this paper, we presented an analytical model for the BTA scheme with ad-hoc scenarios under noisy channel environments. Various simulations are conducted to validate the proposed model. The model is then used to compare the BTA scheme with the legacy DCF, and to determine the optimal block size under different channel conditions.

Appendix: The Markov Chain

In [18], Bianchi first introduced a bi-dimensional stochastic process $\{s(t), b(t)\}$ to model the backoff behavior of the legacy DCF. Process $b(t)$ represents the backoff counter, and it is decremented at the beginning of each slot. For an idle slot, the time scale of $b(t)$ corresponds to a real slot time. In a collision slot, however, $b(t)$ is frozen for the duration of this transmission. Whenever $b(t)$ reaches zero the STA transmits and starts another round of backoff regardless of the outcome of the transmission. The new backoff starts from a value selected randomly from 0 to the current CW size. The CW shall be reset after a successful transmission and be doubled up to a maximum value CW_{max} for corrupted cases. This implies that $b(t)$ de-

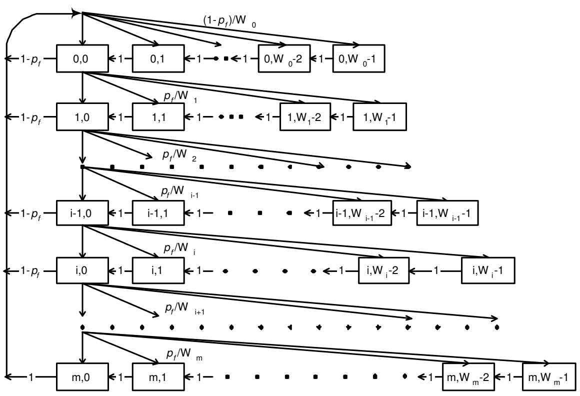


Figure 9: The Markov chain used in this paper

pends on the transmission history, therefore it is a non-Markovian process. To overcome this, another process $s(t)$ is defined to track the CW size.

This bi-dimensional stochastic process is a Markov chain under the following two assumptions. First, the transmission probability τ is constant in every slot time. Second, at each transmission attempt, regardless of the number of retransmission, each frame is collided with an independent constant probability p_f.

Under these assumptions, the bi-dimensional stochastic process $\{s(t), b(t)\}$ forms a Markov chain as shown in Fig. 9. In this chain, all the states are ergodic because they are aperiodic, recurrent and non-null, and thus a stationary solution exists [27]. Given the stationary distribution, we can solve τ and p_f with this Markov chain as follows:

Let us derive the first formula between p_f and τ. In the above Markov chain, p_f stands for the probability that the CW size is doubled because of either collisions or errors. Bianchi's model assumes that there are no transmission errors, so $p_f = p_c = 1-(1-\tau)^{n-1}$, where n stands for the number of STAs in the system. We add the impact of transmission errors in this paper. If the CW is reset after an erroneous transmission, then $p_f = p_c$; if the CW is doubled, and then $p_f = p_c+p_e-p_c{\cdot}p_e$, where p_e is defined in Equation (8). In this paper, we assume that the STA resets the CW size in the case of errors by taking into account that transmission errors occur when one and only one STA is transmitting.

Now, we introduce the second formula between p_f and τ. The transmission probability τ in a slot time should be the sum of all the probabilities of the contention window decreases to zero at all the backoff stages, i.e., $\tau = \sum_{i=0}^{m} b_{i,0}$ where *m* is the maximum backoff stage as defined by $CW_{max} = 2^m \cdot CW_{min}$, and $b_{i,0}$ is the probability of the contention window decreases to zero at the stage i. Bianchi's model assumes that a frame can be retransmitted with infinite times, which is not in accordance with the IEEE 802.11 specification [3]. Wu et al. loose this assumption in their work [19]. We in this paper use formulas similar to [19] and [20] to solve $b_{i,0}$.

Finally, with these two formulas, a closed form solution for p_f and τ is formed and both of them can be solved. Therefore, we can calculate the probabilities in Equations (3-6).

Acknowledgements

The authors wish to thank Dr. Thierry Turletti for many helpful discussions.

References

[1] Kumar S, Raghavan VS, Deng J. Medium access control for ad hoc wireless networks: a survey. *Elsevier Ad Hoc Networks Journal*, 2004.

[2] Jurdak R, Lopes CV, Baldi R. A Survey, Classification and comparative analysis of medium access control protocols for ad hoc networks. *IEEE Communications Surveys*, 1st Quarter 2004.

[3] IEEE 802.11 WG. Part 11: wireless LAN MAC and physical layer specifications. IEEE Std 802.11, 1999.

[4] Xiao Y, Rosdahl J, Throughput and delay limits of IEEE 802.11, IEEE Communications Letters, vol. 6, no. 8, Aug 2002 pp. 355-357.

[5] Xiao Y, Rosdahl J. Performance analysis and enhancement for the current and future IEEE 802.11 MAC protocols. *ACM SIGMOBILE Mobile Computing and Communications Review (MC2R)*, special issue on Wireless Home Networks, Vol. 7, No. 2, Apr. 2003, pp. 6-19.

[6] Ni Q, Aad I, Barakat C, and Turletti T, Modelling and Analysis of Slow CW Decrease for IEEE 802.11 WLAN, Proceedings of IEEE International Symposium on Personnel, Indoor and Mobile Radio

Communications (PIMRC 2003), Beijing, China, September 2003.

[7] Kwon Y, Fang YG, Latchman H. A novel MAC protocol with fast collision resolution for wireless LANs. *IEEE INFOCOM* 2003.

[8] Xiao Y, Li H, Wu K, Leung K, and Ni Q, Reservation with Grouping Stations for the IEEE 802.11 DCF, Proceedings of Networking 2005, pp. 395 - 405, 2005.

[9] Xiao Y, Li H, Wu K, Leung K, and Ni Q, On Optimizing Backoff Counter Reservation and Classifying Stations for the IEEE 802.11 Distributed Wireless LANs, IEEE Transactions on Parallel and Distributed Systems, accepted and to appear.

[10] Tourrilhes J, Packet Frame Grouping: Improving IP multimedia performance over CSMA/CA, Proc. of ICUPC 1998.

[11] IEEE 802.11 WG. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: medium access control (MAC) quality of service (QoS) enhancements. IEEE 802.11e D8.0, February 2004.

[12] Xiao Y, Concatenation and Piggyback Mechanisms for the IEEE 802.11 MAC, Proceedings of IEEE Wireless Communications and Networking Conference 2004, (IEEE WCNC 2004), pp. 1636-1641.

[13] Xiao Y, IEEE 802.11 Performance Enhancement via Concatenation and Piggyback Mechanisms, IEEE Transactions on Wireless Communications, Vol. 4, No. 5, Sep. 2005, pp. 2182- 2192.

[14] Xiao Y, Packing Mechanisms for the IEEE 802.11n Wireless LANs, Proceedings of The IEEE Global Telecommunications Conference 2004 (IEEE GLOBECOM 2004), pp. 7198-7198.

[15] Xiao Y, Efficient MAC Strategies for the IEEE 802.11n Wireless LANs, Journal of Wireless Communications and Mobile Computing, John Wiley and Sons, accepted and to appear.

[16] Ni Q, Li T, Turletti T, and Xiao Y., AFR Partial MAC Proposal for IEEE 802.11n, IEEE 802.11n Working Group Document: IEEE 802.11-04-0950-00-000n, August 13, 2004. 3.

[17] Li T, Ni Q, Malone D, Leith D, Xiao Y, and Turletti T, A New MAC Scheme for Very High-Speed WLANs, Proceedings of IEEE Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2006), accepted.

[18] Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, Vol. 18, Number 3, March 2000.

[19] Wu HT, Peng Y, Long K, Cheng SD, Ma J. Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement. *IEEE INFOCOM* 2002.

[20] Ni Q, Li TJ, Turletti T, Xiao Y. Saturation throughput analysis of error-prone 802.11 wireless networks. *Wiley Journal of Wireless Communications and Mobile Computing (WCMC)*, Vol. 5, Issue 8, pp. 945-956. Dec. 2005.

[21] Tinnirello I, Choi S. Efficiency analysis of burst transmissions with block ACK in contention-based 802.11e WLANs. *IEEE ICC* 2005.

[22] Cover T, Thomas J. *Elements of Information Theory*. John Wiley & Sons, 1991.

[23] Heusse M, Rousseau F, Berger-Sabbatel G, Duda A. Performance anomaly of 802.11b. *IEEE INFOCOM* 2003.

[24] NS-2 simulator. http://www.isi.edu/nsnam/ns/.

[25] IEEE 802.11 WG. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band. IEEE Std. 802.11a, Sept. 1999.

[26] Jain R. *The Art of Computer Systems Performance Analysis: Techniques for Experiment Design, Measurement, Simulation and Modeling.* John Wiley and Sons, Inc, 1991.

[27] Kleinrock L. *Queueing Systems, Volume 1: Theory*. John Wiley & Sons, 1975.

Tianji Li received the B.S. and M.S. degrees in computer science from JiLin and ZhongShan Universities, China, in 1998 and 2001, respectively, and the M.S. degree in networking and distributed computation from the University of Nice Sophia Antipolis, France, in 2004. Currently, he is working towards the Ph.D. degree at the Hamilton Institute, National University of Ireland at Maynooth, Ireland. From 2001 to 2003, he was a software engineer at the Beijing Research Institute of Huawei Technologies, China. His research interests are performance evaluation and optimization in wireless networks. (E-mail: tianji.li@nuim.ie)

Qiang Ni received the B.Eng., M.Sc. and Ph.D. degrees from Huazhong University of Science and Technology (HUST), Wuhan City, China in 1993, 1996 and 1999 respectively. He is currently a faculty member in the Electronic and Computer Engineering Division, School of Engineering and Design, Brunel University, West London, U.K. Between 2004 and 2005 he was a Senior Researcher at the Hamilton Institute, National University of Ireland, Maynooth. From 1999 to 2001, he was a post-doctoral research fellow in the multimedia and wireless communication laboratory, HUST, China. He visited and conducted research at the wireless and networking group of Microsoft Research Asia Lab during the year of 2000. From Sept. 2001 until May 2004, he was a research staff member at the Plante group of INRIA Sophia Antipolis, France. Since 2002, he has been active as a voting member at the IEEE 802.11 wireless LAN standard working group. He has served as Technical Program Committee (TPC) member/session chair for a number of international conferences on wireless communications and networking. His current research interests include communication protocol design, performance analysis, cross-layer optimizations and security issues for wireless networks, and adaptive multimedia transmission over hybrid wired/wireless networks. He has authored/co-authored over 40 international journal/conference papers, book chapters, and standard drafts in this field. He is a member of IEEE. (E-mail: Qiang.Ni@ieee.org)

Yang Xiao worked at Micro Linear as a MAC (Medium Access Control) architect involving the IEEE 802.11 standard enhancement work before he joined Department of Computer Science at The University of Memphis in 2002. Dr. Xiao is an IEEE senior member. He was a voting member of IEEE 802.11 Working Group from 2001 to 2004. He currently serves as Editor-in-Chief for *International Journal of Security and Networks (IJSN)* and for *International Journal of Sensor Networks (IJSNet)*. He serves as an associate editor or on editorial boards for the following refereed journals: *(Wiley) International Journal of Communication Systems*, *(Wiley) Wireless Communications and Mobile Computing (WCMC)*, *EURASIP Journal on Wireless Communications and Networking (WCN)*, and *International Journal of Wireless and Mobile Computing (IJWMC)*. He serves as a guest editor for journal special issues of WCMC, WCN, IJSN, IJWMC, *(Elsevier) Computer Communications*, and *IEEE Wireless Communications*. He serves series co-editor for the book series on "Computer and Network Security," World Scientific Publishing Co. He serves as a referee/reviewer for many funding agencies, as well as a panelist for NSF. He has served as a TPC member for more than 60 conferences including ICDCS, ICC, GLOBECOM, and WCNC. His research areas include wireless networks, mobile computing, and network security. (E-mail: yangxiao@ieee.org)