Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 51 (2007) 1574–1600
Computer Networks
www.elsevier.com/locate/comnet

Achieving optimal performance in IEEE 802.11 wireless LANs with the combination of link adaptation and adaptive backoff ☆

Bo Li [a,*], Roberto Battiti [b]

[a] School of Electronics and Information, Northwestern Polytechnic University, Xian 710072, China
[b] Department of Computer Science and Telecommunications, University of Trento, 38050 POVO Trento, Italy

Received 13 February 2006; received in revised form 7 July 2006; accepted 31 August 2006
Available online 5 October 2006

Responsible Editor: L. Lenzini

Abstract

IEEE 802.11 is one of the most popular wireless LAN standards [Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE Standard 802.11, August 1999]. In the paper, we propose a simple analytical model, which helps one obtain deep insight into the mechanism of achieving optimal performance by using IEEE 802.11 DCF (Distributed Coordination Function) protocol. The first contribution of this paper is the analysis of the optimal operation point where maximum goodput can be achieved. Through the analysis, we answer some fundamental questions about the existence and the uniqueness of the optimal operation point; about the maximum system goodput can be achieved; about the existence of a simple rule to check out if the system operates under the optimal state or not; and how do the data transmission rates, which is dependent on the selected physical transmission mode, and packet transmission errors, caused by channel fading and (or) interference, affect the final system performance. Another contribution is the proposal of a simple distributed adaptive scheme "LABS" (i.e., Link adaptation and Adaptive Backoff Scheme), which makes the system operate under the optimal operation point and, at the same time, achieves some pre-defined target service differentiation ratio between different traffic flows. In the LABS, two adaptive schemes are combined: one is the so called "Link Adaptation" scheme, which dynamically selects an optimal modulation mode at a given time so as to improve the achieved system goodput; the other one is the so called "Adaptive Backoff" scheme, which adaptively adjusts the minimum contention window size of each sending node to guarantee that the system operates under (or near) the optimal operation point. Results obtained in the paper are relevant to both theoretical research and implementation of real systems.

Keywords: Wireless LAN; IEEE 802.11 MAC; Service differentiation; Performance analysis; Link adaptation

☆ The work is supported by the Chinese NSF grant 60572144, and Shaanxi NSF grant 2005F27.
* Corresponding author. Tel.: +86 29 88493556.
E-mail address: libo.npu@nwpu.edu.cn (B. Li).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2006.08.018

1. Introduction

In the literature, performance evaluation of IEEE 802.11 DCF has been executed by using simulation [2] or by means of analytical models [3–8]. One of the core mechanisms for DCF is the backoff procedure. In [3–5], constant or geometrically distributed backoff window sizes have been considered. In [6], an exponential backoff with only two stages is modeled by using a two-dimensional Markov chain. In [7], a more general model that accounts for all the exponential backoff protocol details is proposed. In [8], the model is further extended to the case where packet transmission errors are considered. Instead of using stochastic analysis, in [9], the average value for a variable is used, which results in an approximate but effective analysis.

However, in the above literatures, only one type of traffic is considered. In order to support service differentiation, various enhancements for IEEE 802.11 DCF have been proposed, such as those proposed in [10–13]. In these literatures, the basic idea of achieving service differentiation is to allocate different access parameter set to different traffic flows so as to control the backoff procedures of different traffic flows, which makes it easier for traffic flows with higher priority to access channel resources. Recently, EDCA (Enhanced Distributed Channel Access) is defined in IEEE 802.11e [14]. In EDCA, four backoff entities are supported within one 802.11e node, with each backoff entity corresponding to a particular Access Categories (AC). Service differentiation for each AC is supported by using AC-specific contention parameters, called EDCA parameter set. Arbitration Inter Frame Space (AIFS[AC]), minimum contention window size $CW_{min}[AC]$ are included in EDCA parameter set for each AC. With shorter AIFS[AC] or (and) smaller $CW_{min}[AC]$, the corresponding backoff entity in a 802.11e node has higher priority to access channel resources, which brings about relatively better QoS for the corresponding traffic flows. Moreover, in order to gain deeper insight into the enhanced IEEE 802.11 DCF with service differentiation support, system modeling and performance analysis are proposed, such as those proposed in [15–19].

In order to improve the system performance, apart from enhancing the backoff mechanism defined in IEEE 802.11 DCF, another technique called ''Link Adaptation'', which is the mechanism of selecting one out of multiple available transmission rates at a given time, is studied in literatures, such as those in [20–23]. It is well known that the higher the transmission rate is, the higher the packet error rate is. When the channel state is good, system performance benefit more from improving the transmission rate. Whereas, under the case of bad channel state, it benefits more by lowering the transmission rate, which helps improve the robusticity of the transmission scheme. Therefore, the basic idea of link adaptive schemes is to find out the optimal tradeoff between transmission rate and packet error rate, which brings about optimal system performance over time-varying wireless channels. The AutoRate Fallback (ARF) protocol is used in Lucent Technologies' WaveLAN-II networking devices [20]. In ARF, the transmission rate alternates between 1 and 2 Mbps physical mode, which is based on the idea of keeping track of a timing function and detecting the number of missed ACK frames. In [21], a Receiver-Based Auto-Rate (RBAR) protocol is proposed. The scheme is based on the RTS/CTS (Request-To-Send/Clear-To-Send) mechanism. The basic idea of RBAR is that the receiver detects the channel quality and feeds this information back to the transmitter by using modified CTS frame. The proper transmission rate is chosen after the transmitter received the information about the estimated channel quality. In the above two schemes, detailed analysis, which helps one to obtain deep insight into the system performance is not given. In [22], the authors propose a MPDU (MAC Protocol Data Unit)-based link adaptation scheme based on detailed goodput analysis. The basic idea of the scheme is that the wireless node computes offline a table of physical modes indexed by the system status and each entry of the table is the best physical mode in the sense of maximizing the expected goodput. Although, a detailed analysis is proposed in [22], a more general case where there are a large number of traffic flows sharing the same wireless channel is not considered. It is evident that in order to optimize the system performance, packet collisions between different traffic flows and the corresponding backoff mechanism should be consider. Another problem in [22] is that the obtained analytical results are still complex so that it can be only used in offline manner. In [23], a theoretical model for Medium Access Control (MAC) and physical (PHY) layer protocols in IEEE 802.11 WLANs is proposed, which allows assessing the effect of distinct modulation schemes and channel models on the system performance. However, the proposed performance analysis is not deep enough to find out how to optimize the system performance.

In the paper, based on the theoretical analyses on the optimal operation point, a distributed adaptive scheme LABS is proposed. In the LABS, "link adaptation" and "adaptive backoff" are combined tightly. By using the LABS, maximum goodput and some pre-defined target service differentiation ratio between traffic flows can be achieved. The rest of this paper is organized as follows: in Section 2, to make this paper self-contained, we give a brief description on the analysis model proposed in [7,8,16–18]. In Section 3, optimal operation point and maximum goodput analysis are given. In Section 4, both simulation and numerical results are given to verify the proposed theoretical analysis. In Section 5, basic relationships for achieving maximum goodput and target service differentiation ratios are outlined and verified. Adaptive scheme LABS is proposed in Section 6. Finally, conclusions and future works are summarized in Section 7.

2. Performance analysis

2.1. System modeling

Considering a one-hop IEEE 802.11 WLAN operates in saturation: a fixed number of traffic flows always have packets available for transmission. In the current BSS (Basic Service Set), there are M (>1) sending nodes, with each node bearing only one traffic flow. Let us consider node i ($1 \leqslant i \leqslant M$). It is assumed that the packet transmission error rate over its corresponding wireless channel is $p_{\mathrm{err},i}$ ($0 \leqslant p_{\mathrm{err},i} \leqslant 1$). W_i and m_i are defined as its minimum contention window size and maximum backoff stage, respectively. Let $b_i(t)$ be the stochastic process representing the backoff time counter for the node at time t. And $s_i(t)$ is the stochastic process representing the current backoff stage $(0, 1, \ldots, m_i)$ for the node at time t. Moreover, parameter $p_{\mathrm{c},i}$ is referred to as conditional collision probability, the probability of a collision seen by a packet belonging to node i at the time of its being transmitted over the channel. It should be note that, only the *basic access* mechanism in IEEE 802.11 DCF is analyzed in the paper. The analysis method can be easily extended to the case of *RTS/CTS access* mechanism.

In the following, a two-dimensional discrete-time Markov chain is used to model the behavior of node i. The states are defined as the combinations of two integers $\{s_i(t), b_i(t)\}$. The Markov chain for node i is shown in Fig. 1, where $p_{\mathrm{f},i}$ is defined as the failure probability for a packet sending. Two reasons cause the failure: one is that the packet collides with other packets; Another reason is that some errors occurred over the transmission channel, which is mainly caused by channel fading and (or) interference. Therefore, it can be given as

$$p_{\mathrm{f},i} = p_{\mathrm{c},i} + (1 - p_{\mathrm{c},i})p_{\mathrm{err},i}. \tag{1}$$

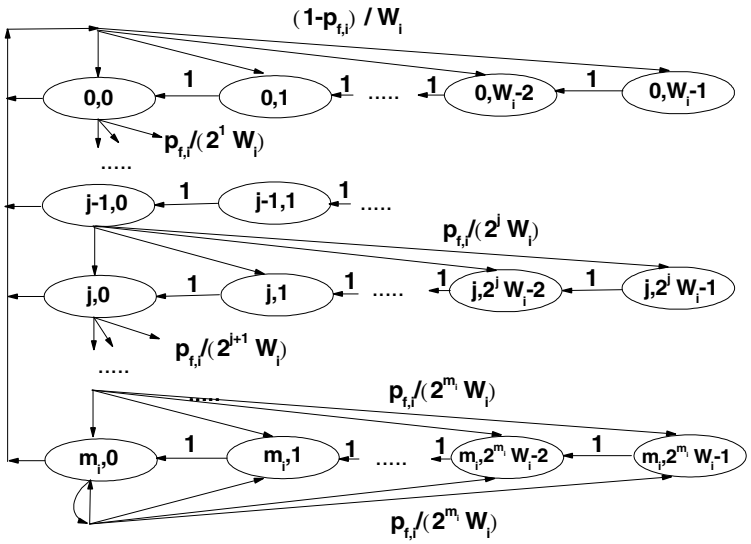


Fig. 1. Markov model for node i.

2.2. Goodput analysis

Let $q_i(j,k)$, $j \in [0,m_i]$ and $k \in [0, 2^j W_i - 1]$, be the stationary distribution of the chain. τ_i is defined as the probability that node i transmits in a randomly chosen slot time. According to [7,8,16–18], we have

$$\tau_i = \sum_{j=0}^{m_i} q_i(j,0) = \frac{2(1-2p_{\mathrm{f},i})}{(1-2p_{\mathrm{f},i})(W_i+1) + p_{\mathrm{f},i}W_i[1-(2p_{\mathrm{f},i})^{m_i}]}. \tag{2}$$

With the above probabilities defined, we can express packet collision probability $p_{\mathrm{c},i}$ as

$$p_{\mathrm{c},i} = 1 - \frac{\prod_{j=1}^{M}(1-\tau_j)}{(1-\tau_i)}. \tag{3}$$

Combining Eqs. (2) and (3) and using SOR (Successive Over Relaxation) numerical method [24], we can get all the values for p_i's and τ_i's.

Random variable c_i is used to indicate whether node i transmits within a randomly selected slot. If it transmits a packet in the selected slot, c_i is set to be 1. Otherwise, c_i is 0. Therefore, we define

$$P(v_1,\ldots,v_M) \triangleq \mathrm{prob}\{c_1 = v_1,\ldots,c_M = v_M\} \quad (v_i = 0 \text{ or } 1,\ 1 \leqslant i \leqslant M). \tag{4}$$

Then, we have

$$P(v_1,\ldots,v_M) = \prod_{i=1}^{M} \tau_i^{v_i}(1-\tau_i)^{1-v_i}. \tag{5}$$

If $\sum_{j=1}^{M} v_j \geqslant 2$, it indicates that more then one nodes transmit in the same selected time slot, which brings about packet collisions.

The average system goodput G, which is a very important performance measurement in this paper, can be defined as

$$\begin{aligned} G &\triangleq \frac{\text{Average useful information bits transmitted in a slot time}}{\text{Average length of a slot time}} \\ &= \sum_{i=1}^{M} g_i = \frac{\sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{\left\{\begin{array}{l} P(v_{j(1\leqslant j\leqslant M)}=0)\cdot\sigma + \sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{s},i} \\ \quad + \sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot p_{\mathrm{err},i}\cdot T_{\mathrm{err},i} \\ \quad + \sum_{\sum_{j=1}^{M} v_j \geqslant 2} P(v_1,\ldots,v_M)\cdot T_{\mathrm{c}}(v_1,\ldots,v_M) \end{array}\right\}} \\ &\triangleq \frac{\sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{\left\{\begin{array}{l} P(v_{j(1\leqslant j\leqslant M)}=0)\cdot\sigma + \sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{s},i} \\ \quad + \sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\cdot p_{\mathrm{err},i}\cdot T_{\mathrm{err},i} \\ \quad + \left[1 - P(v_{j(1\leqslant j\leqslant M)}=0) - \sum_{i=1}^{M} P(v_i=1, v_{j(1\leqslant j\leqslant M, j\neq i)}=0)\right]\cdot T_{\mathrm{c}} \end{array}\right\}}, \end{aligned} \tag{6}$$

where g_i denotes the average goodput contributed by node i. $T_{\mathrm{Len},i}$ is the average time duration for transmitting the useful payload in a data packet sent by node i. R_i (measured in bits per second) is the channel transmission rate for the payload, which is determined, in this paper, by the selected modulation mode ϕ_i for node i. For simplicity, it is assumed that all the packets sent by a node have the same payload transmission duration (i.e., $T_{\mathrm{Len},i}$ is a fixed value). σ is the duration of an empty time slot. $T_{\mathrm{s},i}$ is the average duration of a slot because of a successful transmission of a packet of node i. $T_{\mathrm{err},i}$ is the average duration of a slot because of packet transmission errors occurred over the wireless channel. $T_{\mathrm{s},i}$ and $T_{\mathrm{err},i}$ can be expressed respectively as

$$T_{\mathrm{s},i} = \mathrm{PHY}_{\mathrm{Len}} + \mathrm{MAC}_{\mathrm{Len},i} + T_{\mathrm{Len},i} + \mathrm{SIFS} + \delta + \mathrm{ACK}_{\mathrm{Len},i} + \mathrm{DIFS} + \delta, \tag{7.a}$$

$$T_{\mathrm{err},i} = \mathrm{PHY}_{\mathrm{Len}} + \mathrm{MAC}_{\mathrm{Len},i} + T_{\mathrm{Len},i} + \mathrm{DIFS} + \delta. \tag{7.b}$$

For node i, $\mathrm{PHY}_{\mathrm{Len}} + \mathrm{MAC}_{\mathrm{Len},i}$ is the time duration used to transmit PHY and MAC header information for each packet, and $\mathrm{ACK}_{\mathrm{Len},i}$ is the time duration for transmitting ACK packet. According to the definition of IEEE 802.11 standard, such as those in IEEE 802.11 b and a, both $\mathrm{MAC}_{\mathrm{Len},i}$ and $\mathrm{ACK}_{\mathrm{Len},i}$ vary with the transmission modes adopted by the sending node. Moreover, in the above equation, δ is the average channel propagation delay. And $T_{\mathrm{c}}(v_1,\ldots,v_M)\left(\sum_{i=1}^{M} v_i \geqslant 2\right)$ is the average duration the channel is sensed busy by each node during a collision caused by simultaneous transmissions of different nodes in the current BSS. It can be expressed as

$$T_{\mathrm{c}}(v_1,\ldots,v_M) = \mathrm{PHY}_{\mathrm{Len}} + \max[v_1(T_{\mathrm{Len},1} + \mathrm{MAC}_{\mathrm{Len},1}),\ldots,v_M(T_{\mathrm{Len},M} + \mathrm{MAC}_{\mathrm{Len},M})] + \mathrm{DIFS} + \delta. \tag{8}$$

In Eq. (6), T_{c} is defined as the average duration the channel is sensed busy during a collision. It can be explicitly given as

$$T_{\mathrm{c}} \triangleq \frac{\sum_{\sum_{j=1}^{M} v_j \geqslant 2} P(v_1,\ldots,v_M)\cdot T_{\mathrm{c}}(v_1,\ldots,v_M)}{1 - P(v_{j(1\leqslant j\leqslant M)} = 0) - \sum_{i=1}^{M} P(v_i = 1, v_{j(1\leqslant j\leqslant M, j\neq i)} = 0)}. \tag{9}$$

From Eq. (3), we can easily derive that

$$(1 - p_{\mathrm{c},i})(1 - \tau_i) = \prod_{j=1}^{M}(1 - \tau_j) \quad (1 \leqslant i \leqslant M). \tag{10}$$

When the minimum contention window size $W_i \gg 1$ $(1 \leqslant i \leqslant M)$, transmission probability $\tau_i \ll 1$. Therefore, from Eq. (10), we have the following approximation:

$$p_{\mathrm{c},i} \approx p_{\mathrm{c},j} \quad (i \neq j). \tag{11}$$

From Eq. (6), we have

$$\frac{g_i}{g_j} = \frac{P(v_i = 1, v_{k(1\leqslant k\leqslant M, k\neq i)} = 0)\cdot(1 - p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{P(v_j = 1, v_{k(1\leqslant k\leqslant M, k\neq j)} = 0)\cdot(1 - p_{\mathrm{err},j})\cdot T_{\mathrm{Len},j}\cdot R_j} = \frac{\frac{\tau_i}{(1-\tau_i)}\cdot(1 - p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{\frac{\tau_j}{(1-\tau_j)}\cdot(1 - p_{\mathrm{err},j})\cdot T_{\mathrm{Len},j}\cdot R_j}. \tag{12.a}$$

Defining $\alpha_{g,(i|j)} \triangleq \frac{g_i}{g_j}(1 \leqslant i,j \leqslant M)$ and $\alpha_{\tau,(i|j)} \triangleq \left(\frac{\tau_i}{1-\tau_i}\right)\Big/\left(\frac{\tau_j}{1-\tau_j}\right)(1 \leqslant i,j \leqslant M)$, the above equation can be expressed as follows:

$$\alpha_{g,(i|j)} = \alpha_{\tau,(i|j)}\frac{(1 - p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{(1 - p_{\mathrm{err},j})\cdot T_{\mathrm{Len},j}\cdot R_j}. \tag{12.b}$$

3. Optimal operation point analysis

In this section, we are interested in maximizing total goodput G defined in (6), while *at the same time* ensuring service differentiation, and the hypothesis in this section is that service differentiation is achieved by allocating bandwidth to individual traffic flows so as to satisfy a given target ratio $\alpha'_{g,(i|1)}$, that is

$$\alpha_{g,(i|1)} = \alpha'_{g,(i|1)} \quad (\alpha'_{g,(i|1)} > 0,\ 1 \leqslant i \leqslant M). \tag{13}$$

Moreover, we define $\alpha_{\tau,(i|1)} \triangleq \left(\frac{\tau_i}{1-\tau_i}\right)\Big/\left(\frac{\tau_1}{1-\tau_1}\right)(1 \leqslant i \leqslant M)$, which is one part of $\alpha_{g,(i|1)}$ (see Eq. (12.b)), as the ratio of packet sending rates between node i and node 1. In the following performance analyses, it can be seen that $\alpha_{\tau,(i|1)}$ takes an important role.

In the following, we always assume that packet sending rates for all the active traffic flows in the current BSS satisfy the constraint that $0 < \tau_i < 1$ $(1 \leqslant i \leqslant M)$.

Theorem 1. *Assume that M (>1) traffic flows coexist in the system. In the case that $\frac{\tau_i}{1-\tau_i} = \alpha_{\tau,(i|1)}\cdot\frac{\tau_1}{1-\tau_1}(\alpha_{\tau,(i|1)} > 0, 1 \leqslant i \leqslant M)$, goodput function $G(\tau_1,\ldots,\tau_M)$ defined in Eq. (6) has one and has only one optimal operation point $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})$ $(1 \leqslant i \leqslant M)$ where the maximum goodput can be achieved.*

Proof. Based on Eqs. (5) and (6), we can write the goodput function as follows:

$$G(\tau_1,\ldots,\tau_M)=\frac{\frac{\tau_1}{1-\tau_1}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{\left\{\begin{array}{l}\sigma+\frac{\tau_1}{1-\tau_1}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{s},i}+\frac{\tau_1}{1-\tau_1}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot p_{\mathrm{err},i}\cdot T_{\mathrm{err},i}\\ \quad+\sum_{k=2}^{M}\left(\frac{\tau_1}{1-\tau_1}\right)^k\cdot\sum_{\sum_{l=1}^{M}v_l=k}T_{\mathrm{c}}(v_1,\ldots,v_M)\cdot\prod_{i=1}^{M}\alpha_{\tau,(i|1)}^{v_i}\end{array}\right\}}$$
$$=\frac{F_1\cdot\chi}{\sigma+\sum_{i=1}^{M}H_i\cdot\chi^i}\triangleq\frac{F(\chi)}{H(\chi)},\tag{14}$$

where $\chi\triangleq\frac{\tau_1}{1-\tau_1}$ $(0<\chi<+\infty)$, $F_1=\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i$, $H_1=\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{s},i}+\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot p_{\mathrm{err},i}\cdot T_{\mathrm{err},i}$, and $H_i=\sum_{\sum_{l=1}^{M}v_l=i}T_{\mathrm{c}}(v_1,\ldots,v_M)\cdot\prod_{j=1}^{M}\alpha_{\tau,(j|1)}^{v_j}(2\leqslant i\leqslant M)$. Moreover, F_1 and H_i $(1\leqslant i\leqslant M)$ are constants which are larger than zero. In order to determine the optimal operation point, we study the function $\left(\frac{F(\chi)}{H(\chi)}\right)'$, which can be given as

$$\left(\frac{F(\chi)}{H(\chi)}\right)'=\frac{F(\chi)'H(\chi)-F(\chi)H(\chi)'}{H(\chi)^2}=\frac{F_1\sigma-F_1\sum_{i=2}^{M}(i-1)H_i\chi^i}{H(\chi)^2}.\tag{15}$$

The optimal solution χ^* should satisfy the following equation:

$$\sum_{i=2}^{M}(i-1)\cdot H_i\cdot(\chi^*)^i=\sigma.\tag{16}$$

Because $\sigma>0$ and $\sum_{i=2}^{M}(i-1)H_i\chi^i$ is a monotone increasing function with values ranging from 0 to $+\infty$ when χ varies from 0 to $+\infty$, the optimal solution χ^* must exist and be unique. From Eq. (15), it can be seen that $\left(\frac{F(\chi)}{H(\chi)}\right)'>0$ when $\chi<\chi^*$ and $\left(\frac{F(\chi)}{H(\chi)}\right)'<0$ when $\chi>\chi^*$. Therefore, the goodput function reaches the maximum value when $\frac{\tau_1^*}{1-\tau_1^*}=\chi^*$. It is evident that the optimal solution varies with parameters $\alpha_{\tau,(i|1)}$ $(1\leqslant i\leqslant M)$. Therefore, we denote the optimal solution as $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})$ $(1\leqslant i\leqslant M)$. □

Up to now, the answer for the question about the existence and uniqueness of the optimal operation point is clear. That is, given $\alpha_{\tau,(i|1)}$ $(1\leqslant i\leqslant M)$, there has and only has one optimal operation point, where maximum goodput can be achieved.

By using Eq. (16), the optimal operation point can be solved numerically. However, for the purposes of real implementation and obtaining deeper insight into the system performance, it is necessary to derive more meaningful and concise approximations for the exact formulas. From Eqs. (14) and (16), we have

$$\frac{\tau_1^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})}{1-\tau_1^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})}\leqslant\sqrt{\frac{\sigma}{H_2}}=\sqrt{\frac{\sigma}{\sum_{\sum_{l=1}^{M}v_l=2}T_{\mathrm{c}}(v_1,\ldots,v_M)\cdot\prod_{j=1}^{M}\alpha_{\tau,(j|1)}^{v_j}}}.\tag{17}$$

From the above equation and Eq. (8), it can be seen that, if $M, T_{\mathrm{Len},i}$ $(1\leqslant i\leqslant M)$, are sufficiently large, the optimal operation point $\tau_1^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})\ll 1$ (it is also true for $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})$ $(i>1)$). Therefore, it is reasonable to limit discussions to the case that $\tau_i\ll 1$ $(i=1,\ldots,M)$.

Next we answer the question about where the optimal operation point is and what is the maximum value for the system goodput under the optimal operation point through the following theorem. Moreover, how the system performance is influenced by packet error rates $p_{\mathrm{err},i}$ and transmission rate R_i can also be answered.

Theorem 2. *Assume that M (>1) traffic flows coexist in the system. Assume that $\frac{\tau_i}{1-\tau_i}=\alpha_{\tau,(i|1)}\cdot\frac{\tau_1}{1-\tau_1}$ $(\alpha_{\tau,(i|1)}>0, 1\leqslant i\leqslant M)$. If $M, T_{Len,i}$ $(1\leqslant i\leqslant M)$ are sufficiently large so that the optimal operation point $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})\ll 1$ $(1\leqslant i\leqslant M)$, the optimal operation point can be approximated as*

$$\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}) \approx \frac{\alpha_{\tau,(i|1)}}{\sqrt{\frac{T_c^*}{2}}\cdot\sum_{j=1}^{M}\alpha_{\tau,(j|1)}} \triangleq \tau_{ap,i}^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}), \tag{18}$$

where $T_c^* \triangleq \frac{T_c}{\sigma}$. *Moreover, if* $T_{Len,1} = \cdots = T_{Len,M} = T_{Len}$ *and* M, T_{Len} *are sufficiently large, the corresponding achieved maximum goodput can be approximated as*

$$G_{\max} \approx \frac{T_{\mathrm{Len}}}{T_s + \sigma K + T_c[K(e^{1/K}-1)-1]} \Bigg/ \left(\sum_{i=1}^{M}\frac{\alpha_{i|G}}{R_i\cdot(1-p_{\mathrm{err},i})}\right), \tag{19}$$

where $K \triangleq \sqrt{T_c^*/2}$, $T_s \triangleq \frac{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot T_{s,i}}{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}$ *and* $\alpha_{i|G} \triangleq \frac{g_i}{\sum_{j=1}^{M}g_j} = \frac{\alpha_{g,(i|1)}}{\sum_{j=1}^{M}\alpha_{g,(j|1)}}$ $(i=1,2,\ldots,M)$.

Proof. According to Theorem 1, because the optimal operation point $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}) \ll 1$ $(1 \leqslant i \leqslant M)$, it is reasonable for us to limit the discussion only within the range of $\tau_i \ll 1$ $(1 \leqslant i \leqslant M)$. Moreover, in this case, the relationship $\frac{\tau_i}{1-\tau_i} = \alpha_{\tau,(i|1)}\cdot\frac{\tau_1}{1-\tau_1}$ can be further approximated as $\tau_i \approx \alpha_{\tau,(i|1)}\cdot\tau_1$, which is used in the following derivations.

First, we try to show that T_c defined in Eq. (9) can be approximately regarded as a constant, if we neglect the case that three or more packets collide with each other at the same time. That is

$$T_c \approx \frac{\sum_{\sum_{j=1}^{M}v_j=2}P(v_1,\ldots,v_M)\cdot T_c(v_1,\ldots,v_M)}{\sum_{\sum_{j=1}^{M}v_j=2}P(v_1,\ldots,v_M)} = \frac{\sum_{1\leqslant i\leqslant M,1\leqslant j\leqslant M,i\neq j}\alpha_{\tau,(i|1)}\alpha_{\tau,(j|1)}\cdot T_c(v_1,\ldots,v_M)}{\sum_{1\leqslant i\leqslant M,1\leqslant j\leqslant M,i\neq j}\alpha_{\tau,(i|1)}\alpha_{\tau,(j|1)}}. \tag{20}$$

From the above approximation, it can be seen that once $T_{\mathrm{Len},i}, M, \alpha_{\tau,(i|1)}$ $(i=1,\ldots,M)$ are fixed, T_c can be approximated as a constant, which is independent of τ_i $(1 \leqslant i \leqslant M)$.

Based on the assumption that $\tau_i^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}) \ll 1$ $(i=1,\ldots,M)$ and Eq. (10), goodput function G defined in Eq. (6) can be approximated as

$$G(\tau_1,\ldots,\tau_M) \approx \frac{\tau_1\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{\left\{\begin{array}{l}\sigma+\tau_1\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot(1-p_{\mathrm{err},i1})\cdot T_{s,i}+\tau_1\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot p_{\mathrm{err},i}\cdot T_{\mathrm{err},i}\\ \quad+\left(\frac{1}{1-p_{c,1}}-1-\tau_1\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\right)\cdot T_c\end{array}\right\}} \triangleq \frac{f(\tau_1)}{h(\tau_1)}. \tag{21}$$

Instead of directly finding the optimal operation point by using function $G(\tau_1,\ldots,\tau_M)$, as an approximation, we determine the approximate optimal operation point by using the function $\frac{f(\tau_1)}{h(\tau_1)}$. The optimal solution should satisfy the following condition

$$\frac{f(\tau_1^*)}{f'(\tau_1^*)} = \frac{h(\tau_1^*)}{h'(\tau_1^*)}. \tag{22}$$

Since

$$\frac{dp_{c,1}}{d\tau_1} \approx (1-p_{c,1})\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}. \tag{23}$$

After substituting (23) into (22) and making some further simplifications, one obtains

$$\tau_1^*T_c^*\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)} = (1-p_{c,1})|_{\tau_1=\tau_1^*}\cdot(1-T_c^*)+T_c^* \approx (1-T_c^*)\prod_{i=1}^{M}(1-\alpha_{\tau,(i|1)}\tau_1^*)+T_c^*. \tag{24}$$

Because $(1-\alpha_{\tau,(i|1)}\tau_1^*) \approx (1-\tau_1^*)^{\alpha_{\tau,(i|1)}}$, the above equation can be further approximated as

$$\tau_1^*T_c^*\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)} \approx (1-\tau_1^*)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot(1-T_c^*)+T_c^*. \tag{25}$$

Therefore, $\tau_1^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})$ can be further approximated as

$$\tau_1^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}) \approx \frac{1}{\sqrt{\frac{T_c^*}{2}\cdot\sum_{j=1}^{M}\alpha_{\tau,(j|1)}}} \triangleq \tau_{ap,1}^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)}).$$

Therefore, Eq. (18) is proved.

Next, we evaluate the maximum goodput can be achieved. Under the condition that M, $T_{\mathrm{Len},i}$ $(1 \leqslant i \leqslant M)$ are sufficiently large, and $T_{\mathrm{Len},1} = \cdots = T_{\mathrm{Len},M} = T_{\mathrm{Len}}$, we substitute the approximate optimal solution $\tau_{ap,1}^*(\alpha_{\tau,(1|1)},\ldots,\alpha_{\tau,(M|1)})$ into Eq. (6). Then, we have

$$\begin{aligned} G_{\max} &\approx \frac{\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot\left(1-p_{\mathrm{err},i}\right)\cdot T_{\mathrm{Len},i}\cdot R_i}{\left\{\begin{array}{l}\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sigma+\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}T_{\mathrm{s},i}\\ \quad+\left[1-\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}-\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\right]\cdot T_c\end{array}\right\}} \\ &= \frac{\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot T_{\mathrm{Len}}}{\left\{\begin{array}{l}\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sigma+\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot T_s\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\\ \quad+\left[1-\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}-\tau_{1,ap}^*\cdot\left(1-\tau_{1,ap}^*\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\right]\cdot T_c\end{array}\right\}} \\ &\quad\cdot\left(\frac{\sum_{i=1}^{M}\alpha_{g,(i|1)}}{\sum_{i=1}^{M}\frac{\alpha_{g,(i|1)}}{R_i\cdot(1-p_{\mathrm{err},i})}}\right). \end{aligned} \tag{26}$$

Because

$$\left(1-\frac{1}{K\cdot\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}\right)^{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}} \approx \mathrm{e}^{-1/K}. \tag{27}$$

Then Eq. (26) can be further approximated as

$$G_{\max} \approx \frac{T_{\mathrm{Len}}}{T_s+\sigma K+T_c[K(\mathrm{e}^{1/K}-1)-1]}\Big/\left(\sum_{i=1}^{M}\frac{\alpha_{i|G}}{R_i\cdot(1-p_{\mathrm{err},i})}\right),$$

where $K \triangleq \sqrt{T_c^*/2}$, $T_s \triangleq \frac{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}\cdot T_{s,i}}{\sum_{i=1}^{M}\alpha_{\tau,(i|1)}}$ and $\alpha_{i|G} \triangleq \frac{g_i}{\sum_{j=1}^{M}g_j} = \frac{\alpha_{g,(i|1)}}{\sum_{j=1}^{M}\alpha_{g,(j|1)}}$ $(i=1,2,\ldots,M)$. Therefore, Eq. (19) is proved. □

From Eq. (19), it can be seen that the maximum goodput $G_{\max}$ is actually a function of four parameters: T_{Len}, R_i, $p_{\mathrm{err},i}$ and $\alpha_{i|G}$. Therefore, based on Eq. (19), it can be found that there are several possible ways to improve system performance by increasing possible achieved $G_{\max}$. The first way is to select proper T_{Len}. With the increasing of T_{Len}, larger portion of channel resources can be used to transmit useful information, which has the potential of increasing the system goodput. However, it is well known that with the increasing of T_{Len}, the larger data packet is prone to be corrupted by wireless channel fading and (or) interference more easily, which has the potential of decreasing the system goodput. Hence, T_{Len} should be carefully selected so as to maximize the achieved $G_{\max}$. The second way is to maximize the corresponding product $R_i\cdot(1-p_{\mathrm{err},i})$ for each sending nodes. It is assumed that for each sending node, there are more then one type of transmission modes can be selected. Therefore, in order to improve the achieved $G_{\max}$, the transmission mode, by which $R_i\cdot(1-p_{\mathrm{err},i})$ can be maximized, is chosen. The third way is to select proper bandwidth allocation ratios $\alpha_{i|G}$

($i = 1, 2, \ldots, M$). That is to say, for node i, the larger the product $R_i \cdot (1 - p_{\mathrm{err},i})$ is, the larger the corresponding $\alpha_{i|G}$ should be, which helps to increase the possible achieved $G_{\max}$. The adaptive scheme LABS proposed in this paper is based on the second way.

Deduction 1. Assume that M ($\geqslant 1$) traffic flows coexist in the system. Assume that $\frac{\tau_i}{1-\tau_i} = \alpha_{\tau,(i|1)} \cdot \frac{\tau_1}{1-\tau_1}$ ($\alpha_{\tau,(i|1)} > 0, 1 \leqslant i \leqslant M$). If M, $T_{\mathrm{Len},i}$ ($1 \leqslant i \leqslant M$) are sufficiently large so that the optimal operation point $\tau_i^*(\alpha_{\tau,(1|1)}, \ldots, \alpha_{\tau,(M|1)}) \ll 1$ ($1 \leqslant i \leqslant M$), then the system operates at or close to the optimal operation point if and only if the packet collision rate is approximately equal to $1 - \mathrm{e}^{-1/K}$ ($K \triangleq \sqrt{T_{\mathrm{c}}^*/2}$).

Proof. Assume that the system works at the optimal operation point. By substituting the optimal solution in Eq. (18) into Eq. (10), we have the following approximations:

$$(1 - p_{\mathrm{c},1}^*)(1 - \tau_{ap,1}^*) = \cdots = (1 - p_{\mathrm{c},M}^*)(1 - \tau_{ap,M}^*) = \prod_{i=1}^{M}(1 - \tau_{ap,i}^*) \approx \prod_{i=1}^{M}(1 - \alpha_{\tau,(i|1)} \cdot \tau_{ap,1}^*)$$
$$\approx (1 - \tau_{ap,1}^*)^{\sum_{i=1}^{M} \alpha_{\tau,(i|1)}} \approx \mathrm{e}^{-1/K}. \tag{28}$$

Therefore, the packet collision rate corresponding to the optimal operation point can be expressed as

$$p_{\mathrm{c},i}^* \approx 1 - \mathrm{e}^{-1/K} \quad (i = 1, 2, \ldots, M). \tag{29}$$

Because the function in Eq. (10) is continuous and monotonic, it is easy to arrive at the conclusion that if packet collision rate is approximately equal to $1 - \mathrm{e}^{-1/K}$, the system must operate at or close to the optimal operation point. □

This conclusion is general because it only depends on the parameter T_{c}. Therefore, by comparing the detected packet collision rate with $1 - \mathrm{e}^{-1/K}$, it is possible for one to check out if the system works close to the optimal operation point or not. Deduction 1 is used in the proposed adaptive scheme LABS.

4. Numerical results

In the previous section, we analyze the optimal operation point and its corresponding maximum goodput. In this section, we verify some important approximated results obtained by using both simulation and numerical methods. In this paper, all the discrete events simulations are developed and executed over OPNET Modeler. In our examples, parameters for the system are summarized in Table 1.

Since the maximum goodput analysis is based on the fundamental Eqs. (2), (3), (5) and (6), in the first example, we verify these equations by simulations. In this example, a single-hop system is considered. In it, there are M sending nodes with each of them bearing only one traffic flow. M traffic flows are divided into two groups, with traffic flows belonging to different groups having different QoS requirements. There are n_1 traffic flows, indexed from 1 to n_1, belonging to group one. The minimum contention window sizes for all the group one traffic flows are set to be a fixed value W_{class1}. There are n_2 ($=M - n_1$) traffic flows, indexed from $n_1 + 1$ to M, belonging to the second group, with their minimum contention window sizes being set to be W_{class2}. The payload transmission durations for packets sent by group one and group two nodes are

Table 1
System parameters

PHY header	20 μs
MAC overload	10.25 μs
ACK	25.58 μs
Propagation delay	1 μs
Slot time	9 μs
SIFS	16 μs
DIFS	34 μs

set to be $T_{\text{Len,class1}}$ and $T_{\text{Len,class2}}$, respectively. In the simulation, system goodput G (measured in Mbps) is obtained by varying W_{class1} and W_{class2}. On the other hand, in order to verify Eqs. (2), (3), (5) and (6), goodput G is numerically calculated using the following procedure: First, given W_{class1} and W_{class2}, packet sending rates τ_i $(i = 1, \ldots, M)$ and packet collision rates $p_{c,i}$ $(i = 1, \ldots, M)$ are calculated based on Eqs. (2) and (3); Then, goodput is estimated by substituting the obtained τ_i's and $p_{c,i}$'s into Eq. (6). Goodput obtained by using simulation and numerical ways are compared in Fig. 2. Comparisons show that numerical results accord with simulation results well, especially for the cases that W_{class1} and W_{class2} are not very small, which verifies the fundamental relationships given in Eqs. (2), (3), (5) and (6).

In the second example, we compare the exact optimal operation points τ_1^* numerically obtained from Eq. (6) with the approximated optimal operation points $\tau_{ap,1}^*$ obtained from Eq. (18). In the example, M (M is an even number) sending nodes with each of them bearing one traffic flow are considered. In Fig. 3, comparison results of optimal operation points are shown versus M. It can be seen that good agreements between the exact and the approximate optimal operation points can be achieved if the number of traffic flows M is not too small. Furthermore, comparisons show that good estimation accuracy can be obtained as long as the estimated optimal operation point τ_i^* $(i = 1, 2, \ldots, M)$ are far less than one, which is the condition we based on when deriving the approximate estimation in Eq. (18).

After verifying the accuracy of the estimation for the optimal operation point, we illustrate the accuracy on the evaluated maximum goodput by using approximate results given in Theorem 2. In the numerical experiments, the maximum system goodput $G_{\max}$ is obtained using three different methods. In the first way, $G_{\max}$ is numerical obtained by searching the goodput function G given in Eq. (6). In the second way, $G_{\max}$ is estimated by substituting the approximated optimal operation point τ_i^* $(i = 1, 2, \ldots, M)$, obtained from Eq. (18), into Eq. (6). In the last way, the maximum goodput is estimated by using Eq. (19). Comparison results are shown in Fig. 4. It can be seen that when the packet error rates $p_{\text{err},i}$ are large or packet payload transmission duration T_{Len} are short, the estimation accuracy for the maximum goodput by Eq. (19) becomes slightly poorer, which is caused by the approximation $T_{\text{err},i} \approx T_{\text{s},i}$ made when Eq. (19) is proved. It is evident that with the packet error rates increasing or the decreasing of T_{Len}, estimation errors introduced by making the approximation $T_{\text{err},i} \approx T_{\text{s},i}$ become more evident.

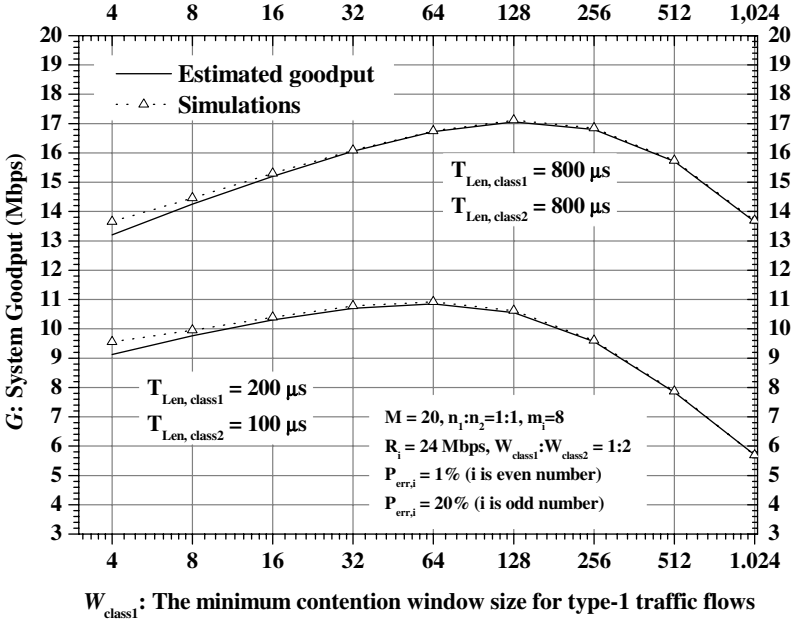


Fig. 2. Verification of fundamental Eqs. (2), (3), (5) and (6).

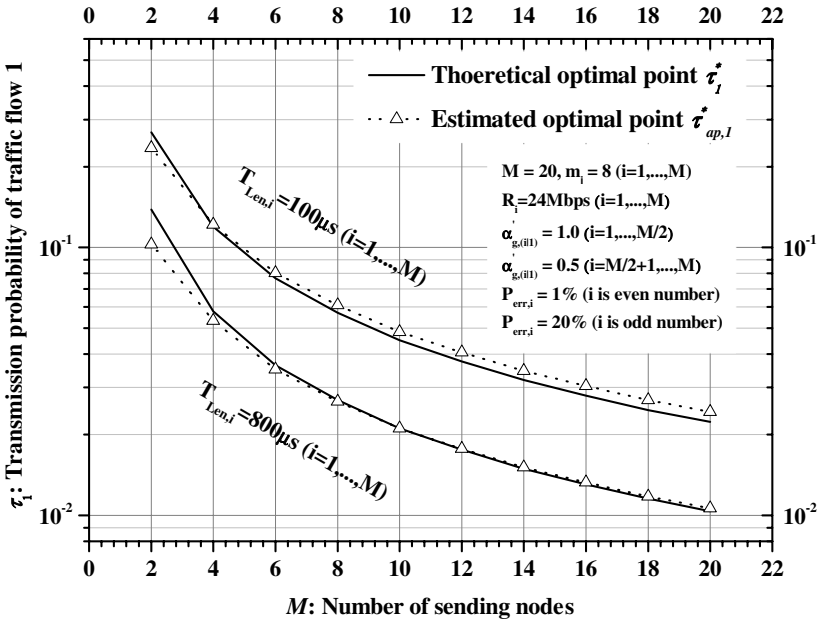


Fig. 3. Comparisons between theoretical optimal operation points and the estimated ones.

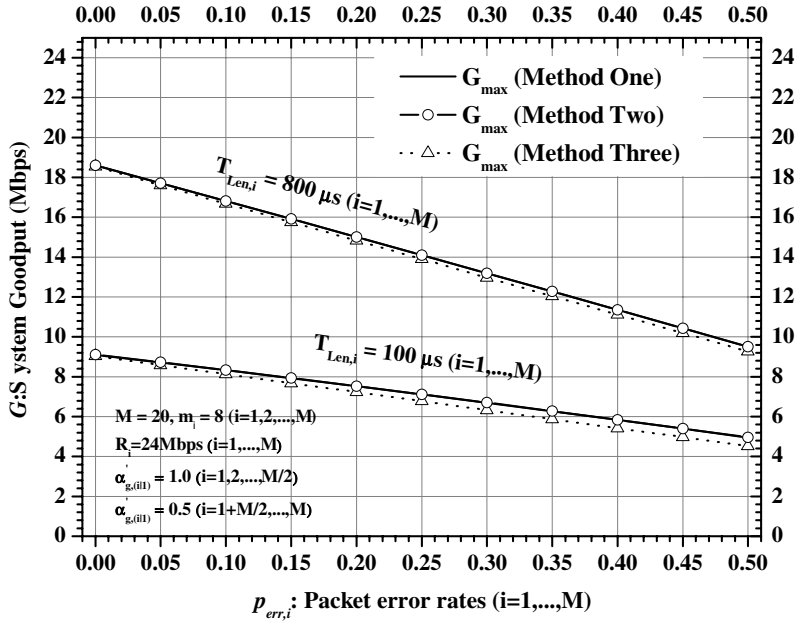


Fig. 4. Estimated maximum goodput $G_{\max}$.

5. Achieving maximum goodput and service differentiation: basic relationships

In this section, basic relationships for achieving maximum system goodput $G_{\max}$ and the given target service differentiation among different traffic flows, that is $\alpha'_{g,(i|1)}$ $(1 \leqslant i \leqslant M)$, are outlined and verified using both simulation and numerical ways, which serves as the basis for designing the adaptive scheme LABS in the next

section. Note that the other way of expressing the given target bandwidth ratios between traffic flows is to use $\alpha'_{i|G}$ $(i = 1, 2, \ldots, M)$ to denote the target value for the corresponding $\alpha_{i|G}$ $(i = 1, 2, \ldots, M)$. It is evident that $\alpha'_{g,(i|1)} = \frac{\alpha'_{i|G}}{\alpha'_{1|G}}$ $(1 \leqslant i \leqslant M)$.

Given the target bandwidth differentiation ratios $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$, based on Eq. (12.b), the ratios for packet sending rates between different traffic flows $\alpha_{\tau,(i|j)}$ can be given as

$$\alpha_{\tau,(i|j)} = \frac{\alpha'_{i|G} \cdot (1 - p_{\mathrm{err},j}) \cdot T_{\mathrm{Len},j} \cdot R_j}{\alpha'_{j|G} \cdot (1 - p_{\mathrm{err},i}) \cdot T_{\mathrm{Len},i} \cdot R_i}. \tag{30}$$

Then the approximated optimal operation point $\tau^*_{ap,i}$, where maximum goodput and target bandwidth differentiations $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$ can be achieved, is obtained by combing Eq. (18) in Theorem 2 with Eq. (30),

$$\tau^*_{ap,i} = \frac{\alpha'_{i|G} \cdot E_j}{\sqrt{\frac{T^*_c}{2}} \cdot (1 - p_{\mathrm{err},i}) \cdot T_{\mathrm{Len},i} \cdot R_i}, \tag{31.a}$$

where E_j $(1 \leqslant j \leqslant M)$ is defined as

$$E_j \triangleq \frac{T_{\mathrm{Len},j} \cdot (1 - p_{\mathrm{err},j}) \cdot R_j}{\alpha'_{j|G} \cdot \sum_{k=1}^{M} \alpha_{\tau,(k|j)}}. \tag{31.b}$$

Note that E_j is used when we describe the LABS.

Based on Eq. (10), the corresponding packet collision rates $p^*_{\mathrm{c},i}(1 \leqslant i \leqslant M)$ at the optimal operation point can be expressed as

$$p^*_{\mathrm{c},i} = 1 - \frac{\prod_{j=1}^{M}(1 - \tau^*_{ap,j})}{(1 - \tau^*_{ap,i})}. \tag{32.a}$$

Moreover, according to the Deduction 1, $p^*_{\mathrm{c},i}$ can be approximately estimated as

$$p^*_{\mathrm{c},i} \approx 1 - \frac{\mathrm{e}^{-1/K}}{(1 - \tau^*_{ap,i})}, \tag{32.b}$$

or being further approximated as (in the case that $\tau^*_{ap,i} \ll 1$)

$$p^*_{\mathrm{c},i} \approx 1 - \mathrm{e}^{-1/K}. \tag{32.c}$$

Finally, based on Eqs. (2) and (3), minimum contention window sizes $W^*_i(1 \leqslant i \leqslant M)$ corresponding to the optimal operation point can be set as follows:

$$W^*_i \approx \frac{2\left(1 - 2p^*_{\mathrm{f},i}\right)}{\tau^*_{ap,i} \cdot \left(1 - 2p^*_{\mathrm{f},i}\right) + \tau^*_{ap,i} \cdot p^*_{\mathrm{f},i} \cdot \left[1 - \left(2p^*_{\mathrm{f},i}\right)^{m_i}\right]}, \tag{33}$$

where according to Eq. (1), $p^*_{\mathrm{f},i} = p^*_{\mathrm{c},i} + \left(1 - p^*_{\mathrm{c},i}\right)p_{\mathrm{err},i}$.

Next, simulation results are given so as to verify equations from (31) to (33). In order to verify the accuracy for relationships proposed in the above three equations, the corresponding W^*_i $(1 \leqslant i \leqslant M)$ are calculated given the target service differentiations ratios $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$ by using equations (31)–(33). Then, in simulations, minimum contention window sizes W_i $(1 \leqslant i \leqslant M)$ for each sending nodes are set to be numerically obtained W^*_i $(1 \leqslant i \leqslant M)$. Finally, the achieved goodput G and bandwidth ratios g_i/g_1 measured from simulations are compared with the corresponding theoretical maximum goodput $G_{\max}$, which can be obtained by numerically searching the maximum value from the function given in Eq. (6). In simulations, target bandwidth differentiation ratios are set as: $\alpha'_{g,(i|1)} = 1.0$ $(1 \leqslant i \leqslant M/2)$ and $\alpha'_{g,(i|1)} = 0.5$ $(1 + M/2 \leqslant i \leqslant M)$.

Fig. 5 shows comparisons between theoretical maximum goodputs and the achieved goodputs measured from simulations by setting $W_i = W^*_i$ $(1 \leqslant i \leqslant M)$. Fig. 6 shows comparisons between the target bandwidth differentiations $\alpha'_{g,(i|1)}$ and the achieved ones g_i/g_1. In these figures, three cases where Eqs. (32.a)–(32.c) are used separately to estimate packet collision rate $p^*_{\mathrm{c},i}$ $(1 \leqslant i \leqslant M)$, are compared. Based on these comparisons, it can

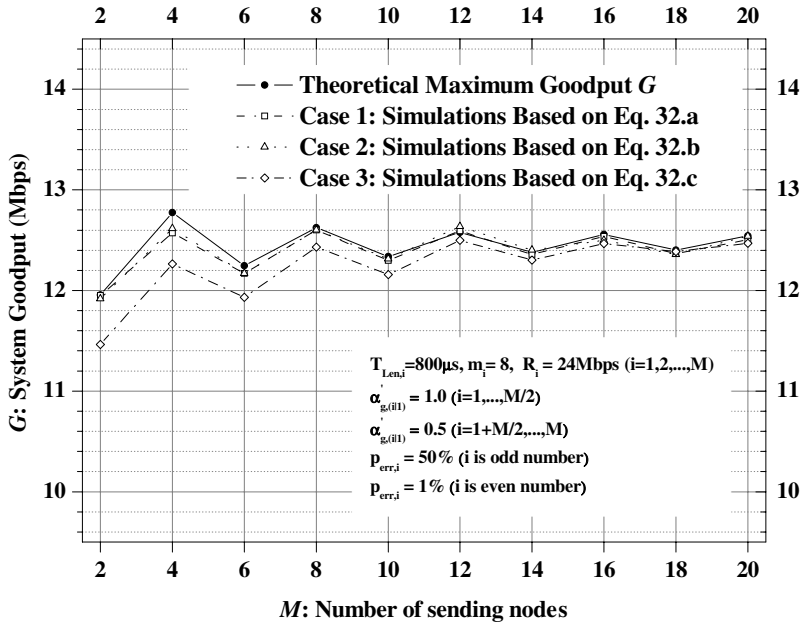


Fig. 5. Comparisons between theoretical maximum goodputs and the achieved goodputs measured from simulations by setting $W_i = W_i^*$ $(1 \leqslant i \leqslant M)$.

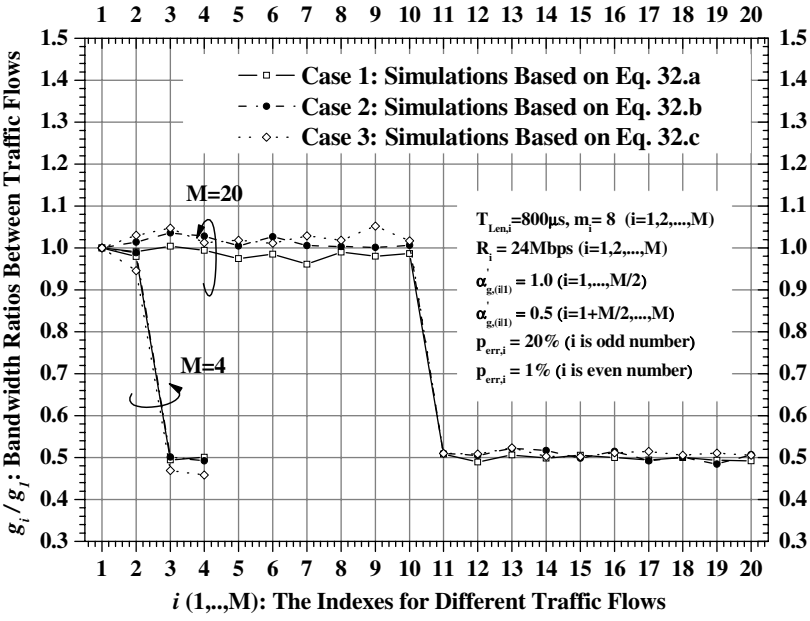


Fig. 6. Comparisons between the target bandwidth differentiations $\alpha'_{g,(i|1)}$ and the achieved ones g_i/g_1.

be found that when the number of traffic flows M is not too small, maximum goodput and target bandwidth differentiations can be achieved in all these three cases. However, in the cases that M is small (say $M < 10$), better performance can be obtained in the first and the second cases. This is because that when M is small, packet sending rates τ_i's cannot be regarded as far less then one anymore. Moreover, it can be seen that

for case 1 and case 2, almost the same performance can be achieved. Therefore, Eq. (32.b), which is simpler, is used to estimate $p^*_{c,i}$ $(1 \leqslant i \leqslant M)$ in the proposed adaptive scheme LABS.

6. The adaptive scheme LABS

In this section, we propose a simple adaptive scheme to achieve maximum goodput and maintain target bandwidth differentiation $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$ between different traffic flows. The system operates under saturation state.

According to equations (12) and (13), bandwidth allocation ratio $\alpha_{g,(i|1)} = \alpha_{\tau,(i|1)} \frac{(1-p_{\mathrm{err},i})\cdot T_{\mathrm{Len},i}\cdot R_i}{(1-p_{\mathrm{err},1})\cdot T_{\mathrm{Len},1}\cdot R_1}$, which indicates that $\alpha_{g,(i|1)}$ can be controlled by $\alpha_{\tau,(i|1)}$ and (or) $T_{\mathrm{Len},i}/T_{\mathrm{Len},1}$. Moreover, $\alpha_{\tau,(i|1)}$ itself can be controlled by adjusting the minimum contention window sizes W_i $(1 \leqslant i \leqslant M)$ for each sending node. On the other hand, $T_{\mathrm{Len},i}/T_{\mathrm{Len},1}$, as for IEEE 802.11e, can be controlled by configuring proper Transmission Opportunity (TXOP) [14]. However, in the following proposed adaptive scheme, it is assumed that packet payloads of all the traffic flows are transmitted in a fixed time duration T_{Len}. Both the optimal operation point and the target bandwidth differentiation ratios $\alpha'_{i|G}(1 \leqslant i \leqslant M)$ are achieved by only adjusting the minimum contention window sizes W_i $(1 \leqslant i \leqslant M)$. To do so, we have the following considerations:

a. With all the traffic flows adopting the same packet payload duration T_{Len}, theoretical results proposed in this paper can be reduced into simpler forms, which helps to simplify the implementations in real-world system. For example, in Theorem 2, it is shown that the goodput can be improved by selecting the optimal modulation mode which can maximize the product $(1 - p_{\mathrm{err},i})\cdot R_i$ in the case of $T_{\mathrm{Len},1} = \cdots = T_{\mathrm{Len},M} = T_{\mathrm{Len}}$.
b. In multi-rate networks with link adaptation supporting mechanisms, such as IEEE 802.11a, in order to improve the achieved goodput, it is helpful to assign transmission time to nodes equally [25]. Within the same transmission time, this allows node with good channel state transmits more data than the node corresponding to bad channel state. Therefore, it is reasonable to assume that $T_{\mathrm{Len},1} = \cdots = T_{\mathrm{Len},M} = T_{\mathrm{Len}}$ in multi-rate scenarios.
c. From Eqs. (30) to (33), it can be seen that by adjusting the minimum contention window sizes W_i $(1 \leqslant i \leqslant M)$, it is easy to achieve both the optimal operation point and the target bandwidth differentiation ratios $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$ at the same time.
d. In IEEE 802.11e, bandwidth differentiations between different traffic flows can also be achieved by allocating different AIFS to different traffic flows. However, it is difficult to find out direct and insight relationships between AIFS and the achieved system performance [18,19], which makes it difficult to use AIFS to achieve accurate service differentiations.

However, by following the same way given in this section, the proposed adaptive scheme can be easily extended to more general cases based on the proposed general theoretical results in this paper. In simulations, it is assumed that $\alpha'_{i|G}$ and T_{Len} are already known by each sending node.

6.1. Basic distributed adaptive scheme

6.1.1. Descriptions of the scheme

Firstly, considering a simpler case where data transmission rates R_i and packet error rates $p_{\mathrm{err},i}$ for all the sending nodes do not vary with time. In order to achieve the maximum goodput and maintain target bandwidth differentiation $\alpha'_{i|G}$ $(1 \leqslant i \leqslant M)$, minimum contention window sizes W_i's should be adaptively adjusted by each sending node. In the following, a basic distributed adaptive scheme is proposed.

In the basic scheme, the function of a sending node i $(1 \leqslant i \leqslant M)$ can be explained as follows: when it is not in the sending state, it listens to the wireless channel to find out if there is some data packet being successfully transmitted by some node. If a data packet sent by node j $(j \neq i)$ is successfully transmitted and captured by node i, the value of variable *Current_E* locally kept by node i is updated by E_j, which is carried in the captured data packet sent by node j. At the same time, node i estimates the value of E_i, which is defined in Eq. (31.b), at

run time. Before sending a data packet, node i inserts the newly estimated E_i into the data packet. If the packet is transmitted successfully, node i modifies its locally kept minimum contention window size $Current\_W_i$ by using the following four equations one by one:

$$\tau^*_{ap,i} = \frac{\alpha'_{i|G} \cdot Current\_E}{\sqrt{\frac{T^*_c}{2}} \cdot (1 - p_{\mathrm{err},i}) \cdot T_{\mathrm{Len},i} \cdot R_i}, \tag{34.a}$$

$$p^*_{\mathrm{c},i} \approx 1 - \frac{\mathrm{e}^{-1/K}}{(1 - \tau^*_{i,ap})}, \tag{34.b}$$

$$W^*_i \approx \frac{2\left(1 - 2p^*_{\mathrm{f},i}\right)}{\tau^*_{ap,i} \cdot \left(1 - 2p^*_{\mathrm{f},i}\right) + \tau^*_{ap,i} \cdot p^*_{\mathrm{f},i} \cdot \left[1 - \left(2p^*_{\mathrm{f},i}\right)^{m_i}\right]}, \tag{34.c}$$

$$Current\_W_i(t+1) = \beta_1 \cdot Current\_W_i(t) + (1 - \beta_1) \cdot W^*_i, \tag{34.d}$$

where $0 < \beta_1 < 1$. Detailed basic adaptive scheme for node i is summarized in Fig. 7.

In order to keep track of the dynamic characteristics of the current BSS, sending nodes monitor the traffic over the common wireless channel and estimates the run-time value of E_i ($1 \leqslant i \leqslant M$). In the following, for node i, we propose a method to estimate E_i ($1 \leqslant i \leqslant M$). In the case that $\tau_j \ll 1$ ($1 \leqslant j \leqslant M$), $\tau_j \approx \alpha_{\tau,(j|i)}\tau_i$ ($1 \leqslant j \leqslant M$). By combining with Eq. (10), we have

$$(1 - p_{\mathrm{c},i}) \approx (1 - \tau_i)^{\sum_{j=1}^{M} \alpha_{\tau,(j|i)}}. \tag{35}$$

Therefore, the estimation of E_i can be given as

$$\widehat{E}_i = \frac{(1 - p_{\mathrm{err},i})R_i}{\alpha'_{i|G} \cdot \frac{\log(1-p_{\mathrm{c},i})}{\log(1-\tau_i)}}. \tag{36}$$

From the above equation, for node i, it can be seen that in order to estimate E_i, three parameters must be estimated. The first parameter is packet error rate $p_{\mathrm{err},i}$. In AWGN (Additive White Gaussian Noise) wireless channel, which is assumed in the paper, $p_{\mathrm{err},i}$ can be directly estimated if both the modulation mode ϕ_i ($i = 1, 2, \ldots, M$) and the corresponding receiver-side SNR (Signal to Noise Ratio) γ_i ($i = 1, 2, \ldots, M$) are known by node i [26]. One way to obtain the estimated SNR is to use the scheme proposed in [21], where the detected receiver-side SNR is feed-backed to the sending node through modified CTS control packet. Another possible way is that the sending node itself estimates the receiver-side SNR directly by estimating the pass loss between the transmitter and the receiver [27]. In this paper, it is assumed that the receiver-side SNR can be perfectly estimated by the sending node. That is, the receiver-side SNR γ_i is known to the transmitter, i.e., node i. Another parameter should be estimated before the estimation of E_i is the packet collision rate $p_{\mathrm{c},i}$. In [28], a way to evaluate the run-time packet collision rate based on an ARMA filter is proposed. Here, we adopt the modified version of this method to estimate the run-time value of packet collision rate. At time t, if a packet has been transmitted successfully or an idle time slot is detected by the considered sending node, the corresponding parameter C_t is set to be 0. C_t is set to be one if the considered sending node detect a busy time slot. If, at time t, a packet transmission failure is seen by the sending node, the corresponding time slot is ignored. This is because that the failure may be caused by packet collisions and (or) packet transmission errors over wireless channel. Therefore, given L ($\geqslant 1$) detected C_{t-j} ($j = 0, \ldots, L-1$) and the current estimation $\hat{p}_{\mathrm{c},i}(t)$ for $p_{\mathrm{c},i}$, we make the following estimation:

$$\hat{p}_{\mathrm{c},i}(t+1) = \alpha_p \cdot \hat{p}_{\mathrm{c},i}(t) + (1 - \alpha_p)\frac{\sum_{j=0}^{L-1} C_{t-j}}{L}, \tag{37}$$

where α_p ($0 < \alpha_p < 1$) is a pre-defined smoothing factor. In Eq. (36), τ_i can be obtained by substituting the estimated $p_{\mathrm{c},i}$ and the current minimum contention window size $Current\_W_i$ into Eq. (2).

For Node i :

if (not in the sending state)

{

Listening to the channel activity, until a data packet sent by station j is received

{

Gets the value of E_j from the captured packet;

Modifies its locally kept variable $Current\_E \Leftarrow E_j$;

}

}

else

{

Estimates $\tau^*_{ap,i} = \dfrac{\alpha'_{i|G} \cdot Current\_E}{\sqrt{\dfrac{T^*_c}{2} \cdot (1 - p_{err,i}) \cdot R_i \cdot T_{Len,i}}}$;

Estimates $p^*_{c,i} \approx 1 - \dfrac{e^{-1/K}}{(1 - \tau^*_{ap,i})}$;

Estimates $p^*_{f,i} = p^*_{c,i} + (1 - p^*_{c,i}) \cdot p_{err,i}$;

Calculates $W^*_i \approx \dfrac{2(1 - 2p^*_{f,i})}{\tau^*_{ap,i} \cdot (1 - 2p^*_{f,i}) + \tau^*_{ap,i} \cdot p^*_{f,i} \cdot [1 - (2p^*_{f,i})^{m_i}]}$;

Estimates $\hat{E}_i \approx \dfrac{(1 - p_{err,i}) \cdot R_i \cdot T_{Len,i}}{\alpha'_{i|G} \cdot \dfrac{\log(1 - p_{c,i})}{\log(1 - \tau_i)}}$;

Smooths the estimated E_i: $E_i(t+1) = \beta_2 \cdot E_i(t) + (1 - \beta_2) \cdot \hat{E}_i$;

Inserts the newly estimated E_i into the data packet;

Sends the data packet;

if (the packet is transmitted successfully)

{

Modifies its $Current\_W_i$ as $Current\_W_i(t+1) = \beta_1 \cdot Current\_W_i(t) + (1 - \beta_1) \cdot W^*_i$;

Modifies its $Current\_E \Leftarrow E_i$;

}

}

Fig. 7. The basic adaptive scheme for node i.

After obtaining $\widehat{E}_i$, E_i can be updated as

$$E_i(t+1) = \beta_2 \cdot E_i(t) + (1 - \beta_2) \cdot \widehat{E}_i, \tag{38}$$

where β_2 ($0 < \beta_2 < 1$) is another pre-defined smoothing factor.

Although basic relationships given in Eqs. (34) and (36) are somewhat complex, it can be found that only two parameters, i.e., $p_{\mathrm{err},i}$ and $p_{\mathrm{c},i}$, should be estimated first. Other parameters, such as τ_i, can be derived based

on the estimated $p_{\mathrm{err},i}$ and (or) $p_{\mathrm{c},i}$. Therefore, the system performance heavily relies on the estimation accuracy for $p_{\mathrm{err},i}$ and $p_{\mathrm{c},i}$. Moreover, compared with the estimation of $p_{\mathrm{c},i}$, the estimation of $p_{\mathrm{err},i}$ are much more challenging. This is because that in order to accurately estimate $p_{\mathrm{err},i}$, the receiver side SNR should be estimated, which turns out to be a rather challenging task because of the highly dynamic characteristics for wireless channels. Moreover, different methods for estimating SNR may lead to different implementation schemes. For example, if SNR is estimated by following the way proposed in [21], feed-back information is needed for the transmitter. No matter what kind of implementation schemes will be adopted, the basic relationships given in equations (34) and (36) do suggest that searching efficient ways in future works to estimate $p_{\mathrm{err},i}$ and $p_{\mathrm{c},i}$ as accurate as possible is helpful and critical for the system performance optimizations.

6.1.2. Performance evaluation

The performance of the basic distributed adaptive scheme is verified by simulations. In simulations, initial values of W_i's for all the sending nodes are set to be 31. In order to compare the performance of the proposed scheme with the theoretical results more easily, channel transmission rates R_i and packet transmission errors $p_{\mathrm{err},i}$ for all the sending nodes are given and do not change with time.

Fig. 8 shows comparisons between theoretical maximum goodputs and the achieved ones measured from simulations by using the proposed scheme. Fig. 9 shows comparisons between the target bandwidth differentiations $\alpha'_{g,(i|1)}$ and the achieved ones g_i/g_1. It can be seen that if the number of traffic flows M is not so small, both maximum goodput and target bandwidth differentiations can be achieved by using the basic adaptive scheme. However, in the cases that M is very small, the basic adaptive scheme achieves worse performance. The first reason is that when M is very small, packet sending rates τ_i's cannot be regarded as far less then one anymore. However, the basic assumptions made for the performance analysis in this paper is that $\tau_i \ll 1$. The second reason is that a lot of parameters are estimated during run-time by the sending nodes themselves. Estimation errors for some parameters increase with the decrease of M. For example, the estimation of E_i depends partially on the estimation accuracy for packet sending rate τ_i. However, the estimation of τ_i is based on the basic relationship given in Eq. (2), which is not very accurate in the case that the number of competing sending nodes is small.

In order to show the convergence process, the variation of the minimum contention window sizes W_3 for node 3 versus time is shown in Fig. 10. As expected, a smaller β_1 and β_2 accelerates the convergence. More

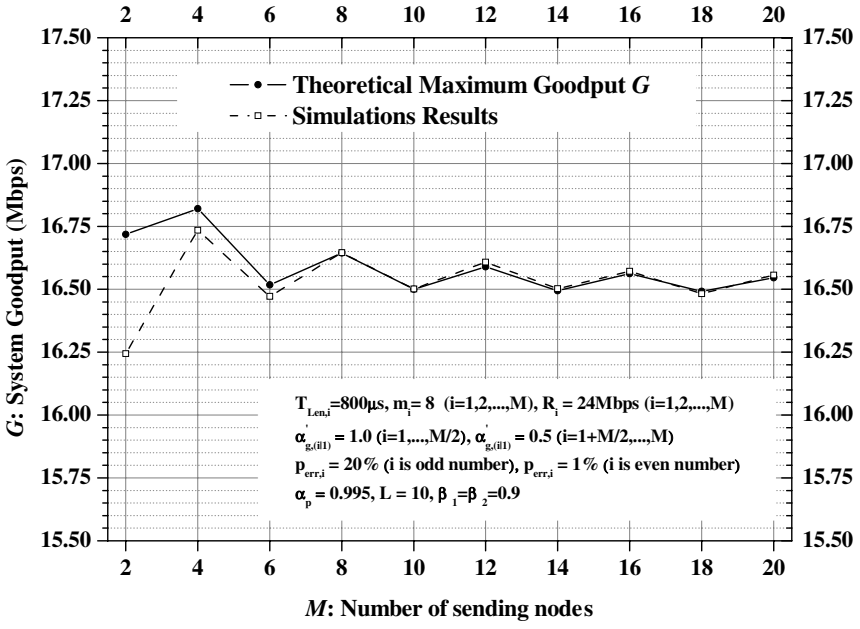


Fig. 8. Goodput achieved by using the basic distributed adaptive scheme.

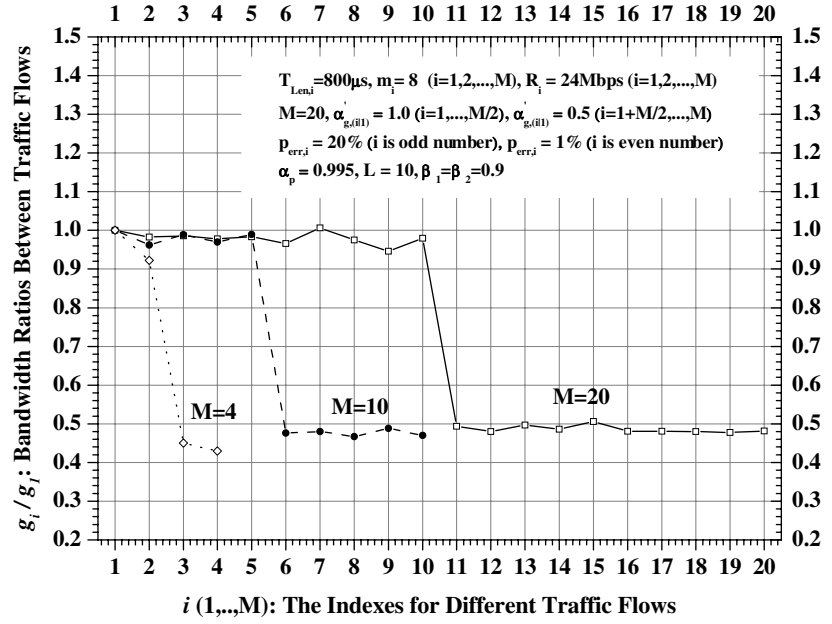


Fig. 9. Service differentiation achieved by using the basic distributed adaptive scheme.

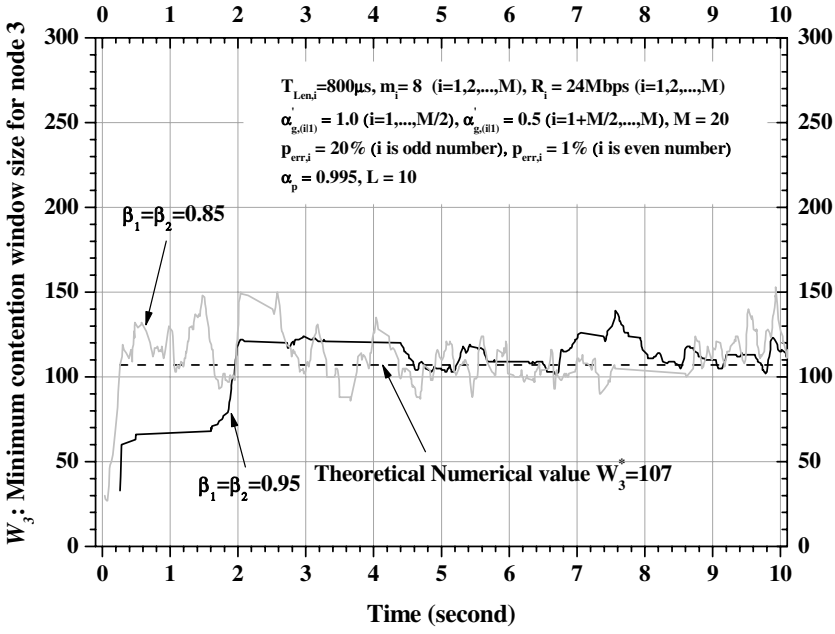


Fig. 10. Convergence speed for the basic distributed adaptive scheme.

detailed analysis in a dynamic context with variable numbers of users will be considered in the future investigations.

Although the maximum goodput $G_{\max}$ can be estimated by using Eq. (19) in Theorem 2, it is difficult to be used in a dynamic network configurations where both channel transmission rates R_i and packet transmission errors $p_{\mathrm{err},i}$ for all the sending nodes change dynamically. In this case, Eq. (29) in Deduction 1, which is more general, can be used to provide ways for mobile nodes to check if the system operates at or near the optimal

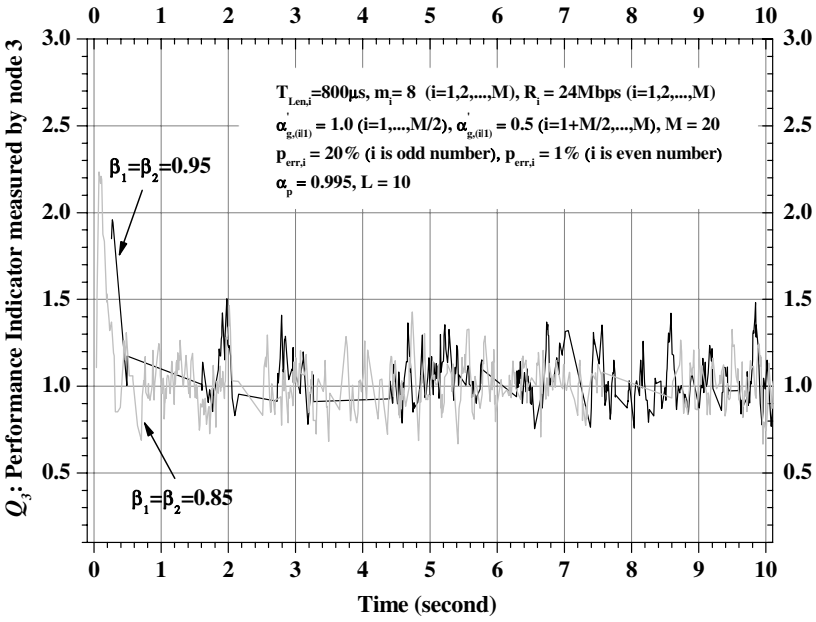


Fig. 11. Performance indicator $\mathbb{Q}_3$ measured by node 3 by using the basic distributed adaptive scheme.

operation point. That is, if the detected value of $\mathbb{Q}_i \triangleq \frac{p_{c,i}}{1-e^{-1/K}} \approx 1$, which is defined as the run-time performance indicator $\mathbb{Q}_i$ measured by node i, then it can be concluded that the system operates at or near the optimal point. Again, the above conclusion holds when the number of active sending nodes is not very small. In Fig. 11, run-time performance indicator $\mathbb{Q}_3$ is shown versus time. It can be seen that the system does converge to the optimal operation point where $\mathbb{Q}_3$ should be approximated by one. Again, it can be seen that a smaller β_1 and β_2 accelerates the convergence.

6.2. Link adaptation and adaptive backoff scheme (LABS)

6.2.1. Descriptions of the scheme

From Eq. (19) in Theorem 2, it can be seen that in order to improve the possible achieved maximum goodput $G_{\max}$, products $R_i\,(1-p_{\mathrm{err},i})$ $(i=1,2,\ldots,M)$ for each sending nodes should be maximized. Under the same channel condition (i.e., the same receiver-side SNR γ_i), $R_i\,(1-p_{\mathrm{err},i})$ is determined by the selected modulation mode ϕ_i. That is to say, if the receiver-side SNR γ_i is given, an optimal modulation mode $\phi_i^*(1 \leqslant \phi_i^* \leqslant \phi_{\max}, 1 \leqslant i \leqslant M)$ should be selected so as to maximize the corresponding $R_i\,(1-p_{\mathrm{err},i})$. For a system without forward error correcting (FEC), Fig. 12 shows the numerical results for $R_i\,(1-p_{\mathrm{err},i})$, according to the assumption of the AWGN wireless channel noise model, versus the receiver-side SNR for different modulation mode selections with packet payload transmission duration $T_{\mathrm{Len},i}=800$ μs. In the paper, for simplicity, only three types of modulation mode is considered, that is $\phi_{\max}=3$. Mode 1 corresponds to QPSK modulation scheme with data rate $R_i(1)=18$ Mbps. Mode 2 corresponds to 16QAM with data rate $R_i(2)=36$ Mbps. Mode 3 corresponds to 64QAM with data rate $R_i(3)=54$ Mbps. From the figure, it can be seen that in order to maximize $R_i\,(1-p_{\mathrm{err},i})$ when the receiver-side SNR $E_{b,i}/N_{0,i}=15$ dB, mode 2 should be selected.

According to the above discussions, if the best modulation mode is selected whenever a sending node i sends a data packet, the former proposed Basic Distributed Adaptive Scheme is extended to the LABS. The best modulation mode is determined as follows:

$$\phi_i^* = \arg\max_{1\leqslant\phi_i\leqslant 3}[R_i(\phi_i)(1-p_{\mathrm{err},i}(\phi_i,\gamma_i))]. \tag{39}$$

The LABS for node i is summarized in Fig. 13.

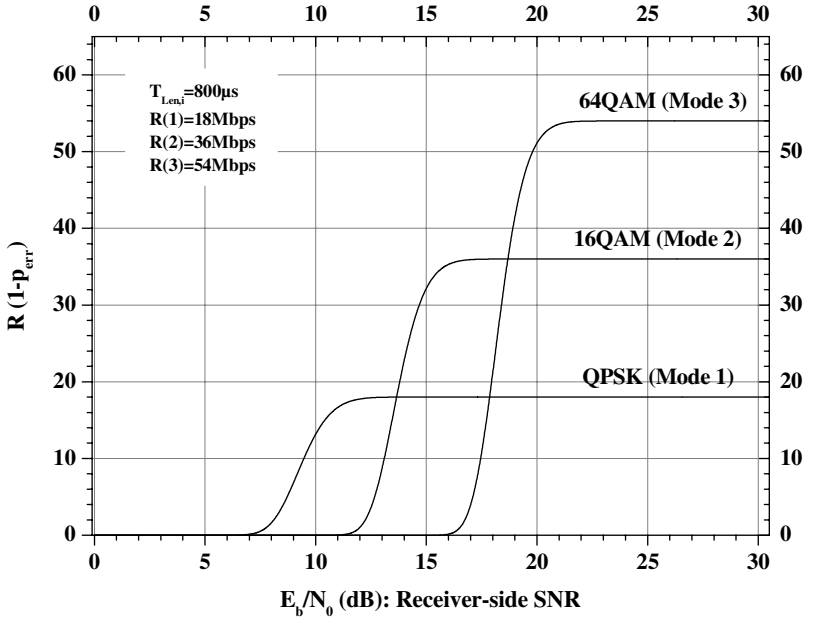


Fig. 12. Numerical results for $R(1-p_{\mathrm{err}})$ versus the receiver-side SNR.

As for the algorithm shown in Fig. 13, we would like to further give some explanations:

a. The modulation selection method given in Eq. (39) can also be explained in another way. Let us assume that the system works at the optimal operation point and $T_{\mathrm{Len},1}=T_{\mathrm{Len},2}=\cdots=T_{\mathrm{Len},M}=T_{\mathrm{Len}}$. In the current time slot, node i will transmit (or re-transmit) a data packet in a time duration of T, which can be approximated as $T\approx T_{\mathrm{Len}}$ as long as T_{Len} is large enough relative to the durations of packet overloads. Therefore, the expected achieved goodput in the coming time slot can be simply given as $\frac{(1-p_{\mathrm{c},i})(1-p_{\mathrm{err},i})\cdot R_i\cdot T_{\mathrm{Len},i}}{T}\approx(1-p_{\mathrm{c},i})(1-p_{\mathrm{err},i})\cdot R_i$, where packet collision rate $p_{\mathrm{c},i}$ can be approximated as a constant value $1-\mathrm{e}^{-1/K}$ (according to Deduction 1) in the case that $T_{\mathrm{Len},1}=T_{\mathrm{Len},2}=\cdots=T_{\mathrm{Len},M}=T_{\mathrm{Len}}$. Therefore, in order to maximize the expected goodput in the coming time slot, it is evident that a best modulation mode ϕ_i^* should be selected so as to maximize the product $(1-p_{\mathrm{err},i})\cdot R_i$, which accord with the conclusion shown in Eqs. (19) and (39).
b. It is assumed that the receiver-side SNR γ_i can be perfectly estimated by the sending node i. However, as to how to estimate the receiver-side SNR efficiently is not the focus of this paper.
c. An AWGN channel noise model is assumed. Therefore, for node i, the symbol error probability for an $M_{\mathrm{mod},i}$-ary ($M_{\mathrm{mod},i}=4,16,64;\ i=1,2,\ldots,M$) QAM with the average SNR per symbol $\gamma_i\left(=\frac{E_{\mathrm{b},i}}{N_{0,i}}\cdot\log_2 M_{\mathrm{mod},i}\right)$ can be given as

$$p_{\mathrm{err\_s},i}(M_{\mathrm{mod},i},\gamma_i)=1-\left[1-2\cdot\left(1-\frac{1}{\sqrt{M_{\mathrm{mod},i}}}\right)\cdot Q\left(\sqrt{\frac{3}{M_{\mathrm{mod},i}-1}\cdot\gamma_i}\right)\right]^2, \tag{40}$$

where the Q-function is defined as

$$Q(x)=\int_x^{\infty}\frac{1}{\sqrt{2\pi}}\cdot\mathrm{e}^{-y^2/2}\mathrm{d}y.$$

For node i, if the packet payload duration $T_{\mathrm{Len},i}$ is not too small, the packet transmission error rate $p_{\mathrm{err},i}(\phi_i,\gamma_i)$ is dominated by the packet payload transmission errors and can be approximated as

$$p_{\mathrm{err},i}(\phi_i,\gamma_i)\approx 1-[1-p_{\mathrm{err\_s},i}(\phi_i,\gamma_i)]^{\frac{T_{\mathrm{Len},i}}{T_{\mathrm{Len\_s},i}}}, \tag{41}$$

For Node i :

if (not in the sending state)

{

Listening to the channel activity, until a data packet sent by station j is received

{

Gets the value of E_j from the captured packet;

Modifies its locally kept variable $Current\_E \Leftarrow E_j$;

}

}

else

{

Estimates the corresponding receiver-side SNR γ_i;

Determines the optimal modulation mode: $\phi_i^* = \arg \max_{1 \le \phi_i \le 3}[R_i(\phi_i)(1 - p_{err,i}(\phi_i, \gamma_i))]$;

Calculates averge transmission rate: $\overline{R_i}(t+1) = \beta_3 \cdot \overline{R_i}(t) + (1-\beta_3) \cdot R_i(\phi_i^*)$;

Calculates averge packet error rate: $\overline{p_{err,i}}(t+1) = \beta_4 \cdot \overline{p_{err,i}}(t) + (1-\beta_4) \cdot p_{err,i}(\phi_i^*, \gamma_i)$;

Estimates $\tau_{ap,i}^* = \dfrac{\alpha'_{i|G} \cdot Current\_E}{\sqrt{\dfrac{T_c^*}{2} \cdot (1 - \overline{p_{err,i}}) \cdot \overline{R_i} \cdot T_{Len,i}}}$;

Estimates $p_{c,i}^* \approx 1 - \dfrac{e^{-1/K}}{(1-\tau_{ap,i}^*)}$;

Estimates $p_{f,i}^* = p_{c,i}^* + (1 - p_{c,i}^*) \cdot \overline{p_{err,i}}$;

Calculates $W_i^* \approx \dfrac{2(1-2p_{f,i}^*)}{\tau_{ap,i}^* \cdot (1-2p_{f,i}^*) + \tau_{ap,i}^* \cdot p_{f,i}^* \cdot [1-(2p_{f,i}^*)^{m_i}]}$;

Estimates $\hat{E}_i \approx \dfrac{(1 - \overline{p_{err,i}}) \cdot \overline{R_i} \cdot T_{Len,i}}{\alpha'_{i|G} \cdot \dfrac{\log(1-p_{c,i})}{\log(1-\tau_i)}}$;

Smooths the estimated E_i: $E_i(t+1) = \beta_2 \cdot E_i(t) + (1-\beta_2) \cdot \hat{E}_i$;

Inserts the newly estimated E_i into the data packet;

Sends the data packet;

if (the packet is transmitted successfully)

{

Modifies its $Current\_W_i$ as $Current\_W_i(t+1) = \beta_1 \cdot Current\_W_i(t) + (1-\beta_1) \cdot W_i^*$;

Modifies its $Current\_E \Leftarrow E_i$;

}

}

Fig. 13. LABS for node *i*.

where $T_{\text{Len\_s},i}$ is the transmission duration of a symbol in the packet payload part for node i. Note that the estimation of packet transmission error rate over more general wireless channel conditions is beyond the discussion of this paper.

d. In order to improve the stability of the algorithm, run-time values for packet transmission rate R_i and packet transmission error rate $p_{\text{err},i}$ are smoothed with parameters β_3 and β_4, respectively.

6.2.2. LABS frame structure

In order to support LABS over IEEE 802.11 MAC protocols, a possible way is to define a new data sub-type frame, which carries an additional LABS field for broadcasting the estimated parameter E_i ($1 \leqslant i \leqslant M$). The frame sub-type is indicated in the frame control field of the MAC header. According to IEEE 802.11 [1], a new frame sub-type can be defined without any problem since there are a number of reserved frame sub-types available in the current 802.11 MAC standard. In Fig. 14, the newly defined data frame structure is shown. The LABS field is 32 bits in length, which carries a floating point value for the estimated E_i ($1 \leqslant i \leqslant M$).

Moreover, as for parameters, such as $\alpha'_{i|G}$ and T_{Len}, they can be broadcasted through beacon frames equipped with new types of Information Elements. Therefore, it can be seen that the proposed LABS can be easily supported by slightly extending the current IEEE 802.11 standard.

6.2.3. Performance evaluation

Based on the same simulation configurations in the Basic Distributed Adaptive Scheme, LABS (see Fig. 13) runs on each sending nodes in stead of the basic distributed adaptive scheme, and, furthermore, a simple two-state discrete time Markov wireless channel is added in the following simulations.

For the two-state channel model, it is assumed that the wireless channel corresponding to the sending node i could be in either a "good" or "bad" state. When the channel is in good state, the corresponding receiver-side SNR $E_{b,i}/N_{0,i}$ ($i = 1, 2, \ldots, M$) is taken from a uniform distribution in the range of 15–30 dB. When in bad state, $E_{b,i}/N_{0,i}$ follows a uniform distribution between 0 and 15 dB. In the channel model, it is assumed that channel state may change even for two adjacent packet retransmissions. p_{gb} is defined as the transition probability for the wireless channel transiting from good state to bad state. p_{bg} is defined as the transition probability for the channel state transiting from bad to good.

For the LABS, link adaptation and adaptive backoff are combined together. As for link adaptation, system goodput is improved by adaptively selecting optimal modulation mode. Whereas, as for the adaptive backoff, system performance is enhanced by guaranteeing the system operates at (or near) the optimal operation point. In the following simulations, we try to show that system performance is further improved by the combination of link adaptation and adaptive backoff. In Fig. 15, performances of four schemes are compared versus different channel conditions (p_{bg}). The first scheme is LABS. And, in the other three schemes, only adaptive backoff is supported. That is, modulation modes for these three schemes are fixed to be 64QAM, 16QAM and QPSK, respectively. Parameters adopted in the simulations are illustrated in the figure. From the figure, it can be seen that system performance is further improved by combining the link adaptation into the adaptive backoff scheme.

In Fig. 16, the LABS is compared with other three link adaptation schemes. In the later three schemes, adaptive backoff is not supported, and the minimum contention window sizes W_i's for all the sending nodes

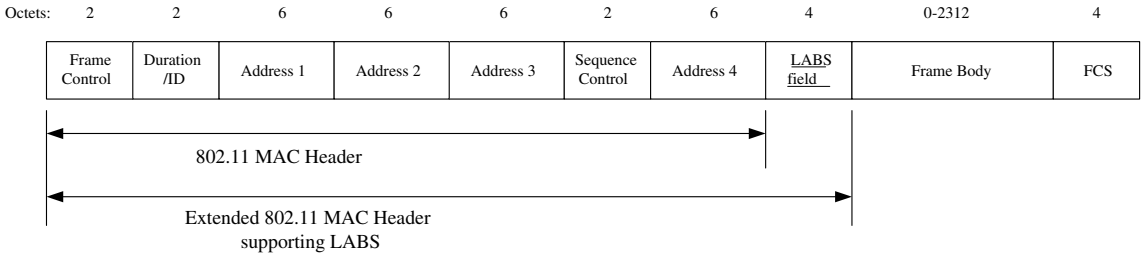


Fig. 14. LABS frame structure.

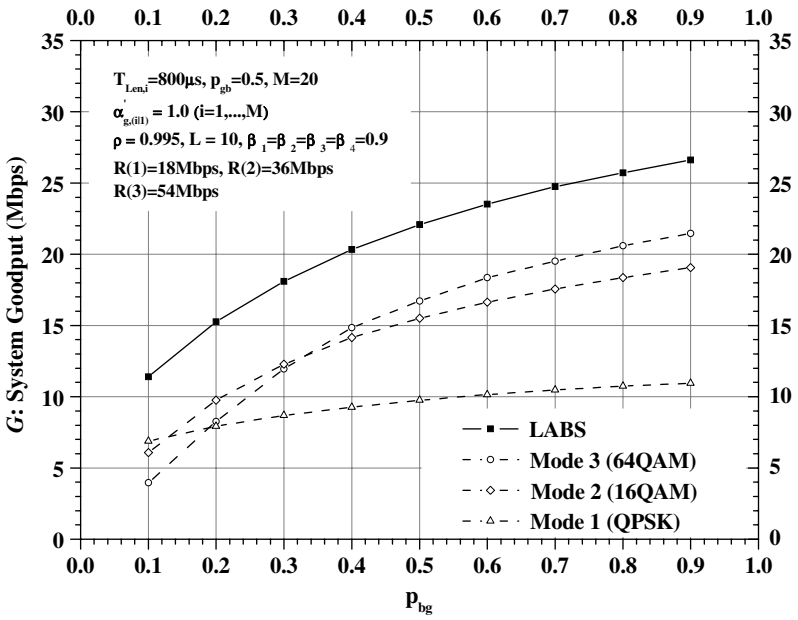


Fig. 15. System goodput G versus channel transition probability p_{bg}.

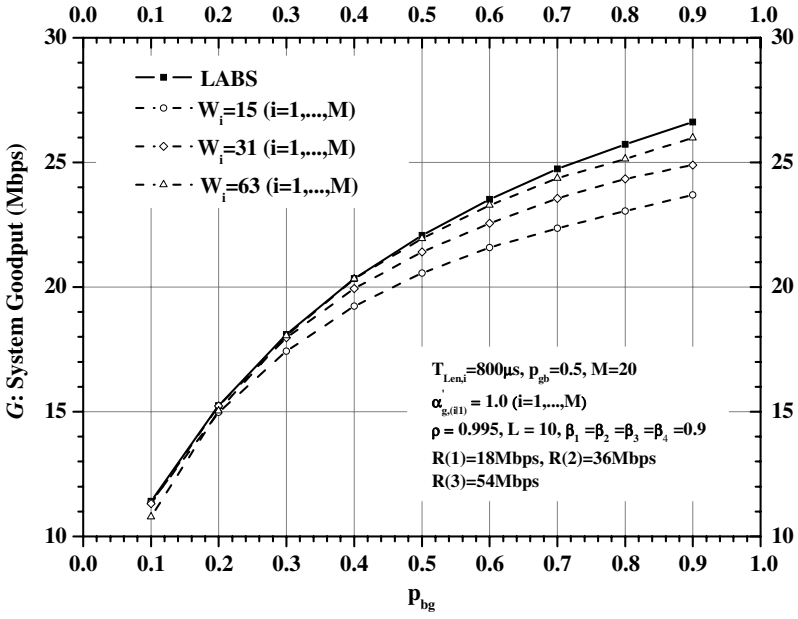


Fig. 16. System goodput G versus channel transition probability p_{bg}.

are set to be 15, 31 and 63, respectively. In the experiments, system goodputs G are obtained by varying channel transition probability p_{bg} from 0.1 to 0.9 (p_{gb} is fixed to be 0.5). From the comparison results, it can be seen that the maximum goodput can be obtained by using the LABS. Moreover, it can be seen that for the schemes without adaptive backoff supporting, their performance come closer to the one of the LABS only at some particular p_{bg}. For example, for the scheme with $W_i = 63$, its performance is close to the one of the LABS when

p_{bg} lies in the range of $0.3 \leqslant p_{bg} \leqslant 0.4$. However, without adaptive backoff supporting, it is impossible to achieve optimal performance under all the channel conditions.

Therefore, can the LABS achieve optimal operation point under all the channel conditions? From the above discussions, we know that performance indicator $\mathbb{Q}_i$ measured by node i can be used to indicate if the system operates at (or near) the optimal operation point. That is, from the point view of node i, if $\mathbb{Q}_i \approx 1$, the system can be regarded as being under the optimal operation point. Here, for simplicity, we define the average system performance indicator $\mathbb{Q} \triangleq \sum_{i=1}^{M} \overline{\mathbb{Q}_i}$, where $\overline{\mathbb{Q}_i}$ is the average value of $\mathbb{Q}_i$ which is measured at run-time. Based on Deduction 1, if $\mathbb{Q} \approx 1$, the system can be regarded as operating under the optimal point. In Fig. 17, system operation points are checked by showing the average system performance indicator $\mathbb{Q}$ versus p_{bg}. It can be seen that the LABS dose achieve optimal operation point under different channel conditions (In the figure, $\mathbb{Q}$ lies between 0.99 and 1.05). Whereas, for the other schemes with fixed minimum contention window sizes W_i, $\mathbb{Q}$ is not always near to 1.0 with the variation of channel conditions. Moreover, by comparing Figs. 16 and 17, it can be seen that performance indicator $\mathbb{Q}$ dose indicate if the system is close to the optimal operation point or not. For example, in Fig. 17, it can be seen that for the scheme with $W_i = 63$, its $\mathbb{Q}$ is closest to 1 in the case of $0.3 \leqslant p_{bg} \leqslant 0.4$. Correspondingly, in Fig. 16, it can be seen that its performance is closest to the one of the LABS in the same range ($0.3 \leqslant p_{bg} \leqslant 0.4$).

In the following experiments, we check if it is possible to achieve the optimal operation point and target bandwidth differentiation $\alpha'_{g,(i|1)}$ at the same time by using the LABS. Fig. 18 shows comparisons between the target bandwidth differentiations $\alpha'_{g,(i|1)}$ and the achieved ones g_i/g_1. In the figure, the achieved bandwidth differentiations in different channel conditions are compared. It can be seen that if channel condition is not too bad (for the case of $p_{bg} = 0.9$ and $p_{bg} = 0.5$), target bandwidth differentiations, that is, $\alpha'_{g,(i|1)} = 1.0$ ($i = 1, \ldots, 10$) and $\alpha'_{g,(i|1)} = 0.5$ ($i = 11, \ldots, 20$), can be achieved by the LABS. However, in the cases that channel condition is worse ($p_{bg} = 0.1$), the LABS achieves worse service differentiation. The reason is that when channel condition is bad, packet sending rates τ_i's are large and cannot be regarded as far less then one anymore.

With the same target bandwidth differentiation requirement, $\mathbb{Q}$ is used to check if the system operates near the optimal operation points or not (see Table 2). In Table 2, the values of $\mathbb{Q}$ under different channel conditions are shown. Once again, it can be seen that the LABS achieves optimal operation point under all the channel conditions ($\mathbb{Q}$ lies in the range from 0.99 to 1.05).

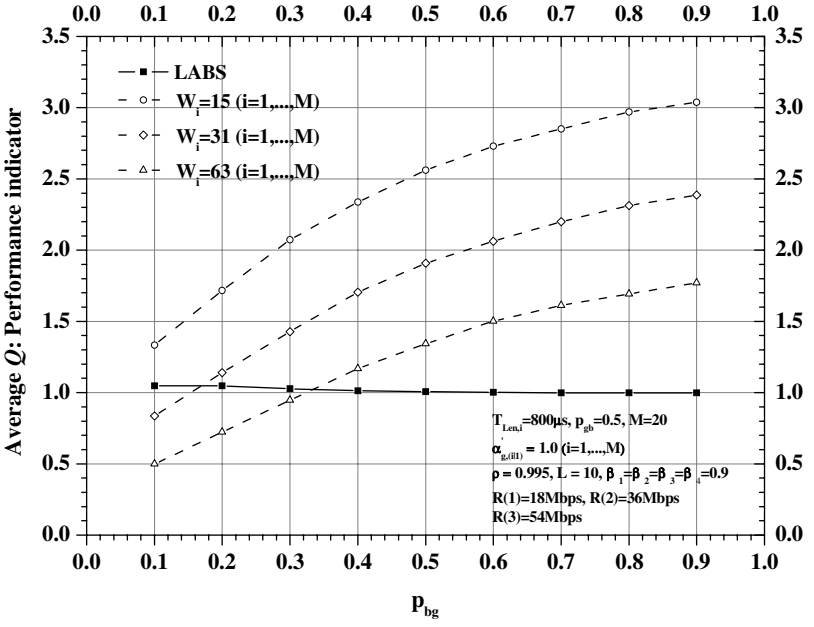


Fig. 17. Average performance indicator $\mathbb{Q}$ versus channel transition probability p_{bg}.

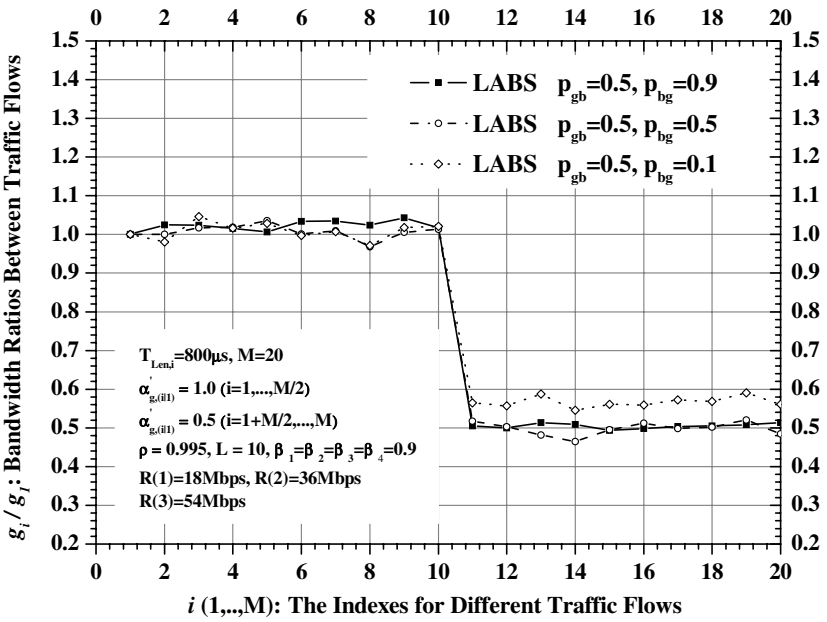


Fig. 18. Service differentiation achieved by the LABS.

Table 2
Performance indicator $\mathbb{Q}$ achieved by the LABS in different channel conditions

p_{bg}	p_{gb}				
	0.1	0.3	0.5	0.7	0.9
0.1	1.02942	1.02906	1.01362	0.99379	0.99147
0.2	1.00320	1.03042	1.04226	1.03949	1.04012
0.3	0.99757	1.01196	1.02302	1.03160	1.03596
0.4	0.99443	1.00202	1.00811	1.01603	1.02293
0.5	0.99393	0.99869	1.00255	1.00706	1.01182
0.6	0.99310	0.99668	0.99868	1.00212	1.00496
0.7	0.99317	0.99474	0.99584	0.99863	1.00122
0.8	0.99259	0.99419	0.99500	0.99674	0.99879
0.9	0.99270	0.99379	0.99421	0.99495	0.99664

$T_{\mathrm{Len},i} = 800$ μs, $M = 20$, $\alpha'_{g,(i|1)} = 1.0$ $(i = 1, \ldots, 10)$ and $\alpha'_{g,(i|1)} = 0.5$ $(i = 11, \ldots, 20)$.
$\rho = 0.995$, $L = 10$, $\beta_i = 0.9$ $(i = 1, 2, 3, 4)$.

7. Conclusions and future work

We propose an analytical model to compute the system goodput. Moreover, based on the model, we derive approximations to get simpler but more meaningful relationships among different parameters. Based on the obtained theoretical results, a distributed adaptive scheme, i.e., LABS, is proposed. In the LABS, ideas of both link adaptation and adaptive backoff are combined, so as to guarantee that the system operate under (or near) the optimal operation point. It is shown that the LABS can achieve the optimal operation point and the target bandwidth differentiation at the same time.

In our future work, we would like to consider how to extend the results obtained in this paper to the following research topics: (1) Non-saturation state should be considered. In this paper, performance analysis is based on the assumption that the system is at the saturation state. Real world system mostly works at non-saturation state. In our recent work [29], it is found that Bianchi's model [7] can be extended to describe

the system performance characteristics in non-saturation state, which suggests that it is possible to combine researches on saturation state and non-saturation state together. (2) In the paper, in order to improve the achieved goodput, the product $R_i(1-p_{\mathrm{err},i})$ is maximized in the proposed LABS. However, according to Eq. (19), it can be seen that there are other ways to further improve the system performance. In our future work, we will consider how to adjust parameters $\alpha_{i|G}$'s, which controls the bandwidth allocations among traffic flows, so as to improve the system performance further. (3) In some more general scenarios, it is possible that not all the sending nodes cooperate to tune their MAC/PHY parameters for optimizing system performance. Considering the case that there are M sending nodes in the current BSS. Among them, nodes 1 to j are legacy IEEE 802.11 nodes, which cannot modify their minimum contention window size W_{Leg}, and nodes $j+1$ to M are enhanced nodes, which are capable of modifying their minimum contention window sizes W_i $(j+1 \leqslant i \leqslant M)$. Based on the proposed theoretical results in this paper, it can be found that by adjusting $W_i (j+1 \leqslant i \leqslant M)$, optimal operation point and the target bandwidth differentiation ratio $\alpha_{g,(i|j+1)}$ $\left(\triangleq g_i/g_{j+1}\right) = \alpha'_{g,(i|j+1)}$ can also be achieved. However, it is evident that, in the future work, the LABS proposed in this paper, which assumes that all the sending nodes in the current BSS can adjust their minimum contention window sizes, should be extended to cope with some more general scenarios where not all the sending nodes wish to cooperate with each other. For achieving this purpose, we think that the first step is to extend the performance analysis proposed in this paper to some more general cases. (4) As for the standardization issues, in this paper, it is assumed that all the nodes can recognize the extended LABS frame (see Fig. 14). However, in the future, the problem of the coexistence of the LABS nodes with legacy IEEE 802.11 nodes, which do not recognize the newly defined LABS frame, should be carefully considered. (5) Performance characteristics of different service differentiation supporting mechanisms should be studied and compared further. In this paper, only one mechanism for supporting service differentiation is studied, i.e., differentiating the minimum contention window sizes according to the priority of different traffic categories. However, in available literatures, such as IEEE 802.11e EDCA, more then one service differentiation supporting mechanisms are proposed. Therefore, advantages and disadvantages of these mechanisms and how these mechanisms can work together are important and interesting research work.

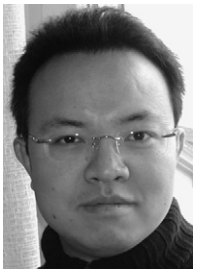
Acknowledgments

We thank the anonymous reviewers for their comments to improve the quality of this paper.

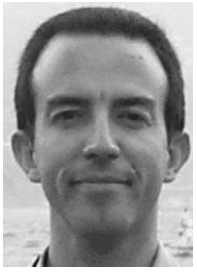
References

[1] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE Standard 802.11, August 1999.
[2] J. Weinmiller, M. Schlager, A. Festag, A. Wolisz, Performance study of access control in wireless LANs IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN, Mobile Networks and Applications 2 (1997) 55–67.
[3] H.S. Chhaya, S. Gupta, Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol, Wireless Networks 3 (1997) 217–234.
[4] T.S. Ho, K.C. Chen, Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LAN's, in: Proceedings of IEEE PIMRC, Taipei, Taiwan, October 1996, pp. 392–396.
[5] F. Cali, M. Conti, E. Gregori, IEEE 802.11 wireless LAN: capacity analysis and protocol enhancement, in: Proceedings of INFOCOM'98, San Francisco, CA, March 1998, vol. 1, pp. 142–149.
[6] G. Bianchi, L. Fratta, M. Oliveri, Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol, in: Proceedings of IEEE PIMRC, Taipei, Taiwan, October 1996, pp. 407–411.
[7] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE Journal on Selected Areas in Communications 18 (3) (2000).
[8] P. Chatzimisios, A.C. Boucouvalas, V. Vitsas, IEEE 802.11 WLANs: performance analysis in presence of bit errors, in: Proceedings of the Communication Systems, Networks and Digital Signal Processing International Symposium (CSNDSP 2004), Newcastle, UK, July 2004, pp. 367–370.
[9] Y.C. Tay, K.C. Chua, A capacity analysis for the IEEE 802.11 MAC Protocol, Wireless Networks 7 (2001) 159–171.
[10] J.L. Sobrinho, A.S. Krishnakumar, Distributed multiple access procedures to provide voice communications over IEEE 802.11 wireless networks, in: Proceedings GLOBECOM 1996, pp. 1689–1694.
[11] J. Deng, R.S. Chang, A priority scheme for IEEE 802.11 DCF access method, IEICE Transactions in Communications 82-B (1) (1999) 96–102.
[12] A. Veres, A.T. Campbell, M. Barry, L.H. Sun, Supporting service differentiation in wirelss packet networks using distributed control, IEEE Journal on Selected Areas in Communications 19 (10) (2001) 2081–2093.
[13] I. Aad, C. Castelluccia, Differentiation mechanisms for IEEE 802.11, in: Proceedings of IEEE Inforcom 2001, pp. 209–218.

[14] IEEE 802.11 WG, IEEE 802.11e/D12.0, Draft supplement to standard for Telecommunications and Information Exchange Between Systems – LAN/MAN specific requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), November 2004.
[15] Z. Jun, G. Zihua, Z. Qian, Z. Wenwu, Performance study of MAC for service differentiation in IEEE 802.11, in: Proceedings of the GLOBECOM, '02, IEEE, vol. 1, November 17–21, 2002, pp. 778–782.
[16] Yang Xiao, A simple and effective priority scheme for IEEE 802.11, IEEE Communications Letters 7 (2) (2003) 70–72.
[17] Bo Li, Roberto Battiti, Performance analysis of an enhanced IEEE 802.11 distributed coordination function supporting service differentiation, QoFIS, in: International Workshop on Quality of Future Internet Services, 2003, Sweden, LNCS 2811, pp. 152–161.
[18] Bo Li, Roberto Battiti, Supporting service differentiation with enhancements of the IEEE 802.11 MAC Protocol: models and analysis, Technical Report of Department of Computer Science and Telecommunications of University of Trento, No. DIT-03-024, available from: <http://eprints.biblio.unitn.it/archive/00000418/>.
[19] Jeffrey W. Robinson, Tejinder S. Randhawa, IEEE saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function, IEEE JSAC 22 (5) (2004) 917–928.
[20] A. Kamerman, L. Monteban, WaveLAN-II: a high-performance wireless LAN for the unlicensed band, Bell Labs Technical Journal (Summer) (1997) 118–133.
[21] G. Holland, N. Vaidya, P. Bahl, A rate-adaptive MAC protocol for multi-hop wireless networks, in: Proceedings of ACM MobiCom'01, July 2001, pp. 236–251.
[22] D. Qiao, S. Choi, K. Shin, Goodput analysis and link adaptation for IEEE 802.11a wireless LANs, IEEE Transactions on Mobile Computing 1 (4) (2002) 278–292.
[23] R. Hoefel, C. Almeida, Performance of IEEE 802.11-based networks with link level adaptive techniques, in: Proceedings of IEEE VTC'2004-fall, vol. 2, pp. 1063–1067.
[24] G. Bolch, S. Greiner, H. de Meer, K.S. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, Wiley-Interscience, 1998, pp. 140–144.
[25] L. Hung, B. Bensaou, Improving wireless local area networks goodput through link adaptation and scheduling, in: IEEE WCNC'2005, pp. 1225–1230.
[26] T.S. Rappaport, Wireless Communications: Principle and Practice, Prentice Hall, Upper Saddle River, NJ, 1996.
[27] D. Qiao, S. Choi, A. Soomro, K.G. Shin, Energy-efficient PCF operation of IEEE 802.11a wireless LAN, Proceedings of IEEE INFOCOM'02 (June) (2002) 580–589.
[28] G. Bianchi, I. Tinnirello, Kalman filter estimation of the number of competing terminals in an IEEE 802.11 Network, in: IEEE INFOCOM 2003, vol. 2, pp. 844–852.
[29] Bo Li, Roberto Battiti, "Analysis of the IEEE 802.11 DCF with service differentiation support in non-saturation conditions", QoFIS, in: International Workshop on Quality of Future Internet Services, 2004, Barcelona, 2004, LNCS 3266, pp. 64–73.



Bo Li received his B.S., M.S. and Ph.D in Communications Engineering from Xidian University of China in 1994, 1996 and 2002, respectively. From 2002 to 2004, he was a Post-doc researcher in the University of Trento. Now he is with Northwestern Polytechnic University in China, where he serve as a full-time professor. His current research interests include wireless multimedia communication networks, resource allocations and cross-layer design.



Roberto Battiti received the Laurea degree from the University of Trento, Italy, in 1985 and the Ph.D. degree from the California Institute of Technology (Caltech), USA, in 1990. He has been a consultant in the area of parallel computing and pattern recognition and since 1991 he has been a faculty member at the University of Trento, where he is now full professor of Computer Networks. His main research interests are heuristic algorithms for optimization problems, in particular reactive search algorithms for maximum clique, maximum satisfiability, graph coloring, networks and massively parallel architectures, code assignment in wireless and cellular networks, protocols for pricing and Quality of Service in wireless networks. He is currently Deputy Dean at the Faculty of Science, representative of the University of Trento in the Management Council of the Italian National Consortium for Computer Science (CINI), member of the advisory committee for the future Telecommunications Plan of the Autonomous Province of Trento, member of the Faculty Commission for Quality Control, evaluator of the Province of Trento for projects related to computer networks and telecommunications. He is the associate editor of various scientific journals and he acted as chair or in the program committee of several international conferences. He is the author of more than 50 scientific publications, including three special issues dedicated to neural computation and experimental algorithmics. He is a member of ACM. IEEE, IEEE Computer Society, and ACM Sigmobile.