Design and Analysis of Cross-Layer Tree Algorithms for Wireless Random Access

Xin Wang, *Member, IEEE,* Yingqun Yu, *Student Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*— In this paper, we develop a random access scheme which combines the widely used binary exponential backoff (BEB) algorithm with a cross-layer tree algorithm (TA) that relies on successive interference cancellation (SIC) with first success (FS). BEB and SICTA/FS complement each other nicely in enabling the novel protocol to attain a maximum stable throughput (MST) as high as 0.6 without packet loss. Although BEB-SICTA/FS avoids the deadlock problem caused by the error propagation commonly present in successive interference cancellation (SIC) algorithms, it may still suffer from deadlock effects induced by the "level skipping" caused by harsh wireless fading effects. We further develop a novel BEB-SICTA/F1 protocol, which is a modified version of BEB-SICTA/FS. Analysis and simulations demonstrate that this simple modification leads to high-throughput random access while completely avoiding deadlock problems.**

***Index Terms*— Random access, successive interference cancellation, tree algorithm, binary exponential backoff.**

I. INTRODUCTION

CONVENTIONAL tree algorithms (TAs) [1]-[3] are designed based on a collision model, where collided packets are discarded. While the so-called first-come-first-serve (FCFS) TA [4] provided 0.487 maximum stable throughput (MST), Tsybakov and Likhanov established that the theoretical MST upper bound of conventional TAs is 0.568 [5]. Inspired by the network-assisted diversity multiple access protocol [6], a so-termed SICTA protocol combining a TA with successive interference cancellation (SIC) was put forth in [7]. By migrating SIC benefits from the physical layer to the TA design, SICTA can afford 0.693 MST which exceeds the 0.568 bound [5] since a number of collided packets are not discarded. However, when applying SICTA to wireless random access, deadlock problems arise due to the error propagation inherent to SIC. To alleviate this problem, we introduced a truncated version of SICTA - SICTA/FS (SICTA with first success) in [8]. SICTA/FS truncates the collision resolution algorithm after the first packet is successfully decoded, and relies on SIC to separate extra packets from the reserved collided packets besides the first success.

In general, a random access protocol consists of two parts: a channel access algorithm and a collision resolution algorithm. Channel access algorithm defines the rules for accessing the physical shared medium, whereas collision resolution algorithm refers to the mechanism used to resolve a collision after it arises [3]. TA-based random access employs a tree approach for collision resolution along with simple channel access algorithms such as gated and window access [3]. We have shown that when equipped with gated access, SICTA/FS can afford 0.6 MST in additive white Gaussian noise (AWGN) channels at the price of a large packet loss. To enable the SICTA/FS gain in a system where minimal packet loss is allowed, we develop here what we term (G)BEB-SICTA/FS protocols where SICTA/FS employs binary exponential backoff (BEB) or gentle BEB(GBEB) algorithms for channel access. BEB has been widely used for channel access in ALOHA protocols [9, Chapter 15]. (G)BEB-SICTA/FS is actually a combination of SICTA/FS and ALOHA. Analysis and simulations reveal that (G)BEB-SICTA/FS can achieve 0.6 MST *without* packet loss.

In SICTA and SICTA/FS, level skipping is employed to enhance the throughput as in a modified TA (MTA) [10]. This level skipping may lead to a protocol deadlock if a single idle slot is incorrectly interpreted as a collision [3]. Although SICTA/FS can avoid the deadlock caused by SIC errors, it may still suffer the deadlock problem if the false alarm of the idle slot is not negligible, which is typical of initial ranging in *fading* channels. To completely avoid the deadlock problem, we modify SICTA/FS and introduce a SICTA/F1 (SICTA with first 1 feedback) protocol, which makes use of binary feedback only: "collision" (e) and "no-collision" (1). Unlike SICTA/FS, SICTA/F1 truncates the collision resolution not only upon a success but also after an idle slot. Then following SICTA/FS, it relies on SIC to separate packets from the reserved collided packets. We further combine SICTA/F1 with BEB or GBEB for use in an IEEE 802.16 setup. Based on analysis and simulations, we establish that besides completely avoiding the deadlock, (G)BEB-SICTA/F1 retains all the SICTA/FS benefits with some MST degradation. Specifically, the MST drops from 0.65 to 0.55 but it is still markedly higher than that of conventional TAs; e.g., [3].

Throughout the paper, we tailor our protocols for the IEEE 802.16 broadband wireless access (BWA) networks [12]. System models are described in Section II. In Sections III and IV,

Manuscript received August 31, 2006; revised February 21, 2007; accepted June 11, 2007. The associate editor coordinating the review of this paper and approving it for publication was R. Negi. Work in this paper was supported in part by the USDoD ARO under Grant W911NF-05-1-0283 and was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

X. Wang is with the Dept. of Electrical Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431 USA (email: xin.wang@fau.edu).

Y. Yu is with Qualcomm Inc., San Diego, USA (email: yyu@qualcomm.com).

Georgios B. Giannakis is with the Dept. of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455 USA (email: georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TWC.2007.060658.

design and analysis of the (G)BEB-SICTA/FS and (G)BEB-SICTA/F1 protocols are presented, respectively. Section VI concludes this paper.

II. SYSTEM SETUP AND MODELING

We consider a finite number of users with infinite size buffers linked wirelessly to a single access point (AP) in an IEEE 802.16 BWA setup [12], adhering to the following modeling conditions:

- *A1) Slotted uplink and scheduling of contention opportunities:* In uplink operation (users to AP), the time is slotted and the physical layer operates in a framed format. The AP uses an uplink map (UL-MAP) message to schedule one Request Information Element (RIE), which consists of a number (n_t) of contention opportunity (CO) slots for random access. Request packets from the users are sent in the CO slots, as in the IEEE 802.16 standards [12].
- *A2) Block 0/1/e (1/e) feedback:* Before each uplink frame, users are informed about the status of CO slots in the last RIE via feedback they receive from the AP, as in IEEE 802.16 [12]. The feedback per slot is one of: a) idle (0) when no packet transmission occurs; b) success (1) upon successful packet reception; or, c) failure (e) upon erroneous packet reception. For SICTA/F1, idleness is also regarded as success; thus a binary (1/e) feedback suffices.
- *A3) Noisy Collision Channel:* The wireless link is a noisy collision channel, where collisions lead to erroneous packet receptions. Moreover, packets can be corrupted by fading and/or noise, even when collisions are absent; see also [8].

To request bandwidth in the IEEE 802.16 [12, Sec. 6.2.8.1], a user transmits a request packet consisting of $L_a = 32$ preamble bits along with a $L_b = 48$ bit bandwidth request message. Quadrature phase shift keying (QPSK) is used to modulate this packet of $L_p = 80$ bits. The packet error rate (PER) for a single non-collided packet depends on whether the channel is modeled as AWGN or fading.

1) AWGN Channels: Let $\gamma := E_b/N_0$ denote the receive signal-to-noise ratio (SNR) per bit, where E_b and N_0 are the bit energy and one-sided noise power density, respectively. For an AWGN channel, the symbol error rate (SER) $P_s^{(0)}$ is given by $P_s^{(0)} = 2Q\left(\sqrt{2\gamma}\right)\left[1 - \frac{1}{2}Q\left(\sqrt{2\gamma}\right)\right]$, where $Q(x) := \int_x^\infty (1/\sqrt{2\pi})e^{-y^2/2}dy$ denotes the Marcum's Q-function [9, Chapter 2]. Assuming that a packet of $L_p/2$ symbols can be successfully recovered only if all its symbols are correctly received, the PER $P_e^{(0)}$ is given by

$$P_e^{(0)} = 1 - (1 - P_s^{(0)})^{L_p/2}. \tag{1}$$

Since a noisy channel is considered, the performance of SIC is no longer perfect. Assuming that the induced noise density N_i per cancellation is also Gaussian distributed, the PER after SIC is similar to $P_e^{(0)}$ in (1), but with reduced SNR. In our ensuing analysis, we assume (for analysis purposes only) that power control is in effect to ensure identical E_b for each user at the AP. Then given that all other collided packets (say n_c) have been correctly decoded, the PER of a packet after SIC is given by [8]

$$P_e^{(n_c)} = 1 - (1 - P_s^{(n_c)})^{L_p/2}, \tag{2}$$

where $P_s^{(n_c)} = 2Q\left(\sqrt{2\gamma'}\right)\left[1 - \frac{1}{2}Q\left(\sqrt{2\gamma'}\right)\right]$, and $\gamma' := E_b/(N_0 + n_c N_i)$.

2) Rayleigh Fading Channels: For independent and identically distributed (i.i.d.) Rayleigh block-fading, the instantaneous PER for uncoded QPSK modulation can be approximated as [14]:

$$P_e^{(0)}(\gamma) \approx \begin{cases} 1, & 0 < \gamma < \gamma_r, \\ a_r e^{-g_r\gamma}, & \gamma \geq \gamma_r \end{cases} \tag{3}$$

where γ is the instantaneous SNR, and a_r, g_r and γ_r are parameters obtained by fitting (3) to the exact PER, as detailed in [14, Appendix]. With $L_p = 80$ bits, we obtain $a_r = 7.3696$, $g_r = 0.5005$ and $\gamma_r = 3.9906$. For Rayleigh fading at average SNR $\bar{\gamma}$, the instantaneous γ is described by the probability density function (pdf): $p_\gamma(\gamma) = 1/\bar{\gamma}e^{-\gamma/\bar{\gamma}}$. Therefore, the average PER is given by

$$\begin{aligned} P_e^{(0)} &= \int_0^{\gamma_r} p_\gamma(\gamma)d\gamma + \int_{\gamma_r}^\infty a_r e^{-g_r\gamma}p_\gamma(\gamma)d\gamma \\ &= 1 - e^{-\gamma_r/\bar{\gamma}} + \frac{a_r}{\bar{\gamma}}\frac{e^{-b_r\gamma_r}}{b_r} \end{aligned} \tag{4}$$

where $b_r := (1/\bar{\gamma}) + g_r$. Assuming the channel fading is independently and identically distributed per slot, we use least-squares criterion to estimate the fading coefficients for the decoded packets in the previously stored collided packets $\mathbf{y}$; i.e, for decoded packet $\mathbf{x}_i$, we obtain its fading coefficient $\hat{h}_i = \arg\min_{h_i} \|\mathbf{y} - h_i\mathbf{x}_i\|^2$. The reconstructed signal $\hat{h}_i\mathbf{x}_i$ is then subtracted from $\mathbf{y}$ using SIC. Again modeling the induced noise per reconstruction/cancellation as Gaussian, the average PER $P_e^{(n_c)}$ after SIC is given accordingly by taking into account the reduced average SNR $\bar{\gamma}'$ as in (2).

In fading channels, even with power control in effect, we still need to consider the packet detection problem. We assume that request packet detection relies on the output of a filter matched to the preamble signal. Whenever two or more users send packets over the same slot, collision occurs and is assumed to be detected with probability 1. However, when there is no packet transmission, false alarm occurs with probability P_F; whereas if a single packet is transmitted per slot, it is detected with probability P_D. With a prescribed threshold T and perfect synchronization assumed among users and with the AP, we have [6]

$$\begin{aligned} P_F &= \exp\left(-\frac{T^2}{2L_aN_0}\right), \\ P_D &= \exp\left(-\frac{T^2}{2L_aN_0(1 + L_a\bar{\gamma})}\right). \end{aligned} \tag{5}$$

Having clarified the system model, we proceed to introduce a simple IEEE 802.16 system, where each RIE only contains $n_t = 1$ CO slot. Per *A1)* and *A2)*, this means that feedback is available at the end of each slot. Extensions to systems with $n_t > 1$ will be discussed in Sec. V.

III. BEB-SICTA/FS AND GBEB-SICTA/FS OVER AWGN CHANNELS

Being a truncated version of the SICTA protocol [7], SICTA/FS shares with SICTA the SIC approach to take advantage of collided packets in a TA, but introduces the idea of first success to effect robustness in wireless random access [15], [8]. In SICTA/FS, we define as collision resolution interval (CRI) the interval of successive slots starting from the slot where an initial collision occurred up to and including the slot at the end of which a success (1) feedback is sent by the AP. In a CRI, SICTA/FS follows the underlying tree structure of SICTA until it arrives at its first success. At that point, a single packet is decoded and SIC is employed to extract as many extra packets from reserved previously collided packets. However, even though some packets may remain unresolved, SICTA/FS terminates the CRI at this first success point and starts another CRI. In so doing, SICTA/FS endows SICTA with three advantages: 1) avoidance of the deadlock problem caused by the error propagation of SIC; 2) robustness to feedback errors; and 3) limited-sensing capability whereby users do not need to monitor the channel history all the time but can easily drop in and out. The performance of SICTA/FS over AWGN channels was studied in [8]. Analysis and simulations revealed that equipped with a gated channel access algorithm, SICTA/FS can afford 0.6 MST. For these reasons, SICTA/FS overcomes limitations of SICTA while offering a practically feasible protocol for wireless networking. However, in SICTA/FS with gated channel access algorithm (denoted as gated-SICTA/FS), high MST comes at the price of 20% packet loss.

A. BEB-SICTA/FS

To capture the SICTA/FS gain in a system where minimal packet loss is allowed, we put forth a protocol that we call BEB-SICTA/FS because it augments SICTA/FS with a BEB-based channel access algorithm. BEB has been widely used for collision avoidance in ALOHA protocols [9, Chapter 15]. In fact, BEB-SICTA/FS can be viewed as a combination of SICTA/FS with ALOHA. The motivation behind this combination is twofold: 1) With a BEB-based channel access algorithm, we can sufficiently reduce the initial collision size and thus enhance the efficiency of SICTA/FS, since the tree truncation in SICTA/FS heavily degrades the throughput when the number of initially collided packets in a CRI is large; 2) By using the practically feasible SICTA/FS scheme to separate the collided packets (instead of simply discarding them as in ALOHA), we can increase throughput. Since ALOHA (BEB) is the standard random access protocol in 802.16 [12], we can easily deploy our BEB-SICTA/FS in such systems as follows.

Whenever a user has a request packet to send, it activates its backoff procedure and generates a random backoff counter (BC) according to its backoff window (BW); see also [12]. This BC is reduced by 1 upon each success (1) feedback, and remains frozen when idle (0) or failure (e) feedback messages are received (i.e., when SICTA/FS is in effect to separate the collision); and then resumes running-down the BC as soon as a success feedback is received. Notice that an idle slot following a success (1) feedback indicates a slot with "zero collided packet", and will lead to a success feedback by the AP. Each user is permitted to transmit when its BC reaches zero. If the feedback for this transmission is success, the request is completed and the user resets its BW to a minimum value BW_{min}. However, if a failure (e) feedback is received, the user doubles its BW as long as it stays smaller than a maximum value BW_{max}, and switches from the backoff procedure to a TA mode; then it follows the tree structure of SICTA/FS, specified by the feedback messages from the AP, until a success feedback is received. If the request packet of a user is separated at the end of the CRI, the user should repeat the contention for this packet with the new BW.

Following this procedure, we can describe BEB-SICTA/FS algorithmically. In the protocol, each user needs to maintain two *local* counters B_t and D_t. In particular, B_t represents its BC value and D_t represents the TA stack level value, which indicates how many slots the user needs to defer in the CRI according to the underlying tree structure, at the beginning of the tth CO slot. Whenever an *active* user has a request packet to send, it randomly selects a B_t according to its BW and sets its $D_t = 0$. At the beginning of each slot t, the user transmits a packet if and only if both its $B_t = 0$ and $D_t = 0$. The update of the local counter values and the corresponding operations at each active user then follow the rules listed below:

1) When feedback = 1,
    a) if $D_{t-1} > 0$, then $D_t = 0$; and if the user has no bandwidth assignment from the UL-MAP and thereby knows its request packet has not been separated by SICTA/FS, a new B_t is randomly selected according to the current BW;
    b) if $B_{t-1} > 0$, then $B_t = B_{t-1} - 1$.
2) When feedback = e and $B_{t-1} > 0$, then $B_t = B_{t-1}$.
3) When feedback = e and $B_{t-1} = 0$,
    a) if last feedback = 1, its BW is doubled as long as it is smaller than the BW_{max};
    b) if $D_{t-1} > 0$, then $D_t = D_{t-1} + 1$;
    c) if $D_{t-1} = 0$, then

$$D_t = \begin{cases} 0, & \text{with probability } p; \\ 1, & \text{with probability } 1-p. \end{cases}$$

4) When feedback = 0 and $B_{t-1} > 0$, then $B_t = B_{t-1}$.
5) When feedback = 0 and $B_{t-1} = 0$,
    a) if $D_{t-1} > 1$, then $D_t = D_{t-1} + 1$;
    b) if $D_{t-1} = 1$, then

$$D_t = \begin{cases} 0, & \text{with probability } p; \\ 1, & \text{with probability } 1-p. \end{cases}$$

B. Saturation Throughput of BEB-SICTA/FS in AWGN channels

In BEB-SICTA/FS, the BEB algorithm is used to coordinate access of the channel. It is well known that queuing analysis of a large system coordinated by BEB is infeasible due to the interaction among multiple queues. To bypass this impasse, Bianchi developed an extremely simple model that accounts for all details of the BEB algorithm at saturation; namely, when the transmission queue of each user in the network is always nonempty [16]. Based on the *saturation throughput*, we can estimate the MST of BEB-SICTA/FS. Here we adopt a framework similar to [16] and capitalize on our performance analysis for SICTA/FS in [8], to investigate the saturation throughput of BEB-SICTA/FS in AWGN channels.

Let l_n and s_n denote the conditional length and conditional number of succeeded packets in a CRI, given that n packets initially collide, and $\overline{L}_n := E[l_n]$ and $\overline{S}_n := E[s_n]$ their expected values. Define $B^i_{n,p}$ as the probability mass at the value i of a binomial random variable with total n trials and success probability p; i.e., $B^i_{n,p} := \binom{n}{i} p^i(1-p)^{n-i}$. From [8], we borrow the following lemma.

Lemma 1: *For AWGN channels, the expected CRI length* $\overline{L}_n$ *and the expected number of successfully decoded packets* $\overline{S}_n$ *in SICTA/FS can be recursively obtained as*

$$\overline{L}_0 = 1; \qquad \overline{L}_1 = \frac{1-P_e^{(0)}+P_e^{(0)}/p}{1-P_e^{(0)}}; \tag{6}$$

$$\overline{L}_n = \frac{1+\sum_{i=1}^{n-1} B^i_{n,p}\overline{L}_i}{1-p^n-(1-p)^n}, \quad \textit{for } n \geq 2; \tag{7}$$

$$\overline{S}_0 = 0; \qquad \overline{S}_1 = 1; \tag{8}$$

$$\overline{S}_n = \frac{\sum_{i=1}^{n-1} B^i_{n,p}\overline{S}_i + B^{n-1}_{n,p}P_s^{SIC}}{1-p^n-(1-p)^n}, \quad \textit{for } n \geq 2; \tag{9}$$

where P_s^{SIC} *denotes the probability that an extra packet can be successfully decoded using SIC, given by*

$$P_s^{SIC} = (1-P_e^{(n-1)})\prod_{m=2}^{n-1}\frac{B^{m-1}_{m,p}}{1-p^m-(1-p)^m}(1-P_e^{(m-1)}); \tag{10}$$

and PER $P_e^{(0)}$ *and* $P_e^{(n_c)}$, $n_c \in [1, n-1]$, *are given by (1) and (2), respectively.*

Let us now consider the beginning instants of the CRIs (denoted by t_c), and rely on a bi-dimensional Markov chain to describe the backoff counter value $b(t_c)$ of one user in BEB as in [16]. Regardless of the number of retransmissions, we assume that each packet collides with constant probability p_c. We further let τ denote the probability of a user transmitting at t_c regardless of its backoff stage; W denote BW_{min} and m denote the “maximum backoff stage” which satisfies $BW_{max} = 2^m W$. Based on these definitions, we have established that (see Appendix for the proof):

Lemma 2: *For a saturation system with* N *users,* p_c *and* τ *are the solutions to the equations*

$$\tau = \frac{2(1-2p_c)}{(1-2p_c)(W+1)+p_cW(1-(2p_c)^m)}, \qquad p_c = 1-(1-\tau)^{N-1}(1-P_e^{(0)}). \tag{11}$$

Note that as described by the algorithm rules in Sec. II-A, we “ignore” the effect of TA in the exponential backoff. Recall that BEB is introduced to reduce the initial collision size for the efficiency of SICTA/FS. To this end, even though a certain user’s request packet is resolved from the collision by TA, this user should still use an increased BW for its new request. With the CRI statistics (6)-(9) at hand and τ calculated from (11), we are ready to estimate the MST of BEB-SICTA/FS as follows.

TABLE I

System parameters used in the IEEE 802.16 simulator

Symbol rate	20 Mbaud
Frame length	1 ms
Minislot length	400 ns (8 symbols)
CO slot length	8 minislots
Request packet length (preamble + BW request)	5 minislots
Request Backoff Start (BW_{min})	4
Request Backoff End (BW_{max})	512
maximum backoff stage m	7
Request Retry (retry limit)	7

Proposition 1: *For a network with* N *users, the MST of BEB-SICTA/FS can be estimated from its saturation throughput as*

$$R = \frac{\sum_{n=0}^{N} P_t(n)\overline{S}_n}{\sum_{n=0}^{N} P_t(n)\overline{L}_n} = \frac{\sum_{n=0}^{N} B^n_{N,\tau}\overline{S}_n}{\sum_{n=0}^{N} B^n_{N,\tau}\overline{L}_n}, \tag{12}$$

where $P_t(n) = B^n_{N,\tau}$ *denotes the probability of* n *users transmitting at the beginning of a CRI.*

Proof: With n initially collided packets, the average number of decoded packets and the average CRI length are given by $\overline{S}_n$ and $\overline{L}_n$, respectively. Considering the probability $P_t(n)$ of n users transmitting at the beginning of a CRI, (12) follows readily. ☐

To validate our analysis, we simulated a 802.16 BWA system with N users, each having a buffer capable of storing up to 5,000 request packets. QPSK is used to modulate the $L_p = 80$ bits of each request packet. The system operates in AWGN with one-sided power density N_0. At most 3 previous erroneous receptions can be stored at the AP. Moreover, SIC is imperfect and induces (Gaussian) noise with variance $N_i = 0.1N_0$. For SICTA/FS, binary splitting is used with $p = 0.5$ (our analytical results suggest selecting $p = 0.58$). The system parameters closely follow those specified in IEEE 802.16 [12] and are listed in Table I. Each simulation is obtained by averaging 10 independent runs, where in each run the simulated system was in operation for 10 seconds.

We test the BEB-SICTA/FS protocol on the simulated system at saturation condition for an AWGN channel with two SNR values: ∞ dB and 6 dB. Per SNR value, 10 cases with different number of users N are tested. In Fig. 1 we first compare analytical with simulated results with the minimum BW size $W = 4$, where henceforth the “lines” are obtained through the derived analytical expressions, while each point depicts the corresponding simulation result. The slight difference between analytical and simulated results is due to the fact that the Markov model is less accurate when W is small [16]. To confirm this, we perform the tests with a larger $W = 16$. As shown in Fig. 1, analytical and simulated results match well. We next compare the saturation throughput of ALOHA (BEB), gated-SICTA/FS and BEB-SICTA/FS. As corroborated by Fig. 2, BEB-SCITA/FS clearly outperforms ALOHA and gated-SICTA/FS in saturation throughput. By allowing packet loss, gated-SICTA/FS can achieve 0.6 MST [8]. However, if packet loss is not allowed, the MST of gated-SICTA/FS is determined by the saturation throughput, which decreases as the number of users increases, as in Fig. 2. By combining

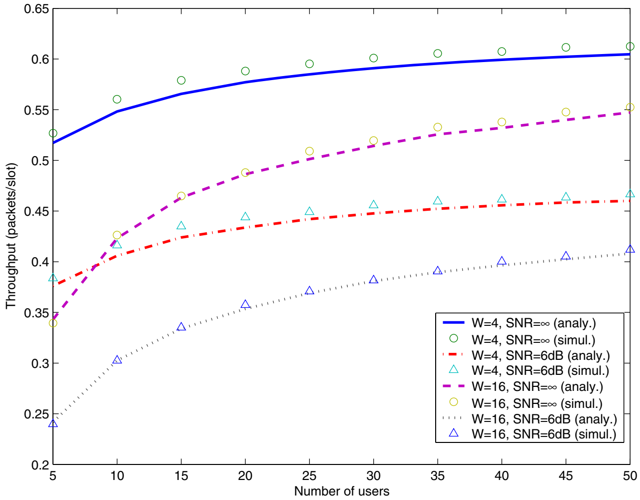


Fig. 1. Comparison between analytical and simulated results for the saturation throughput of BEB-SICTA/FS over AWGN channels.

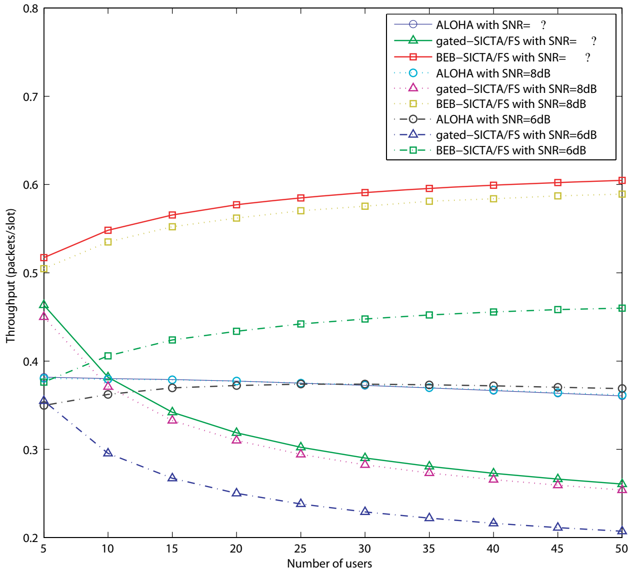


Fig. 2. Saturation throughput comparison among ALOHA (BEB), gated-SICTA/FS and BEB-SICTA/FS over AWGN channels.

SICTA/FS with BEB, we have been able to achieve 0.6 MST *without* packet loss as the number of users increases.

C. GBEB-SICTA/FS

In the IEEE 802.16 BWA system, the AP can take a more active role in random access; e.g., the AP may frequently update a BW value in the Uplink Channel Descriptor (UCD) message so that all users can use an identical and hopefully optimal BW [12, Sec. 6.2.8]. Here we resort to the saturation throughput analysis to assess the benefit of this common BW scheme. If the same BW W_i, $i \in [0, m]$, is employed by N users, the probability τ that a user transmits in a given slot, turns out to be independent of the collision probability p_c and is given by [c.f. (11)]: $\tau = 2/(W_i + 1)$. Using this τ and Proposition 1, we can obtain the system saturation throughput $R_g(W_i, N)$ for each W_i and N. An intuitively appealing scheme (denoted as CBEB-SICTA/FS) is given by $W_{opt}(N) = \arg\{\max_{i=0,\cdots,m} R_g(W_i, N)\}$. The saturation throughput attained by this scheme is given by $R_g^{opt}(N) = R_g(W_{opt}(N), N)$. Fig. 3 compares the saturation throughput between CBEB-SICTA/FS and BEB-SICTA/FS. It is seen that the former provides similar saturation throughput over all N, and outperforms the latter, especially when the number of users is small.

Since the number of users in random access is difficult to estimate and changes from time to time, the CBEB-SICTA/FS may not be practically feasible. To approximate it, we propose a modified protocol that we term GBEB-SICTA/FS because it relies on a gentle BEB (GBEB)-based channel access algorithm. The GBEB is motivated by features similar to the CSMA/CCA protocol in [17]. In GBEB, the AP has a prescribed BW change threshold V_{th} and maintains two counters n_f and n_s for BW updates, where n_f and n_s represent the number of consecutive collision CRIs and the number of consecutive non-idle single-slot CRIs, respectively. At the end of each CRI, the AP updates n_f, n_s and the common BW as follows: 1) If the CRI contains $l > 1$ CO slots, the AP sets $n_s = 0$ and $n_f = n_f + 1$. If $n_f \geq V_{th}$, the AP sets $n_f = 0$ and doubles its BW when it is less than BW_{max}; and 2) If the CRI contains $l = 1$ non-idle CO slot, the AP sets $n_f = 0$ and $n_s = n_s + 1$. If $n_s \geq V_{th}$, the AP sets $n_s = 0$ and halves its BW when it is greater than BW_{min}.

Upon receiving an updated BW different from the available one, each deferring user in the last CRI needs to update its BC value $b(t_c + 1)$ from the old value $b(t_c)$ using [17]

$$\begin{cases} b(t_c+1) = 2b(t_c) + \lfloor 2x_{rnd} \rfloor, & \text{if BW is doubled,} \\ b(t_c+1) = \lfloor b(t_c)/2 \rfloor, & \text{if BW is halved.} \end{cases} \tag{13}$$

where x_{rnd} denotes a random real number uniformly distributed in [0,1) and $\lfloor \cdot \rfloor$ stands for the floor operation. By adding these BW updates into the rules of BEB-SICTA/FS, we obtain the GBEB-SICTA/FS protocol. Fig. 3 also compares the saturation throughput between GBEB-SICTA/FS with threshold $V_{th} = 20$ and BEB-SICTA/FS. It is seen that the GBEB-SICTA/FS provides similar saturation throughput over all N and outperforms BEB-SICTA/FS, when the number of users is small. The advantage of GBEB-SICTA/FS over BEB-SICTA/FS becomes noticeable for $N \leq 25$.

IV. BEB-SICTA/F1 AND GBEB-SICTA/F1 OVER RAYLEIGH FADING CHANNELS

Level skipping is used by MTA to skip a subsequent slot when a collision slot is followed by an idle slot [3], [11]. SICTA/FS can take advantage of level skipping to attain high throughput. In fading channels, however, the false alarm probability P_F of an idleness can become non-negligible. As a result, erroneous level skipping could lead SICTA/FS to a deadlock of "perpetual splitting". For instance, consider two packets collide in the initial slot. According to TA, each user involved tosses a two-sided coin and joins either the right sub-tree (to transmit) with probability p, or, the left sub-tree (to defer) with probability $1 - p$. If the spitting results in both users joining the left sub-tree, an idle slot arises next and may be falsely detected as a "collision" with non-negligible probability P_F. The TA then requires the non-existing users

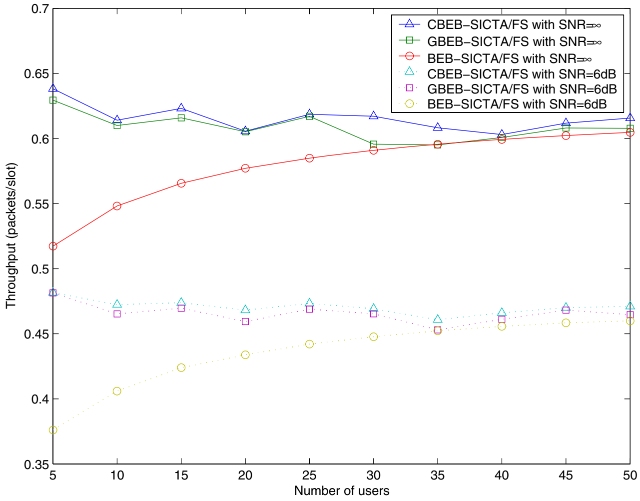


Fig. 3. Saturation throughput comparison between BEB-SICTA/FS and GBEB-SICTA/FS and CBEB-SICTA/FS over AWGN channels.

to split, which would surely give rise to idle subsequent slots. Level skipping then renders perpetual splitting during the CRI, until an external criterion terminates it; see also [3], [11]. To resolve this deadlock, an easy remedy is to skip levels only a finite and predetermined number of times in succession [3]. In the proposed SICTA/F1 protocol, we resort to a more aggressive solution which simply terminates the CRI whenever an *idleness* or a success is encountered. This alternative enables very simple operation and binary feedback. That is, the feedback per slot is now one of: a) collision (e) upon erroneous packet reception; or b) no-collision (1) otherwise. Clearly, while SICTA/FS is a truncated version of SICTA, SICTA/F1 can be seen as a truncated version of SICTA/FS. Although truncation lowers the MST by a certain amount, analysis and simulations will show that SICTA/F1 can still achieve markedly high throughput while being simple and robust to fading effects. To appreciate the simplicity of SICTA/F1 in operation, we next describe it algorithmically. In SICTA/F1, each user only needs to maintain a *local binary* counter D_t, which is set to 0 or 1 at the beginning of a CRI depending on whether the user has a request packet to send or not. At the beginning of each slot t, each user transmits its packet if and only if $D_t = 0$. With p denoting the splitting probability of a TA, the update of D_t per user entails the following simple rules:

1) If feedback = e and $D_t = 1$, then $D_{t+1} = 1$.
2) If feedback = e and $D_t = 0$, then
$$D_{t+1} = \begin{cases} 0, & \text{with probability } p; \\ 1, & \text{with probability } 1-p. \end{cases}$$
3) If feedback = 1, the user is assured that the current CRI ends and a new CRI begins afterwards. If the user has been assigned bandwidth from the UL-MAP and thereby knows its request has succeeded, it removes the head-of-line packet from its buffer.

A. *CRI Statistics of SICTA/F1*

Again, let l_n and s_n denote the length and number of successful packets in a CRI of SICTA/F1, given that n packets initially collide. We wish to derive their expected values $\overline{L}_n := E[l_n]$ and $\overline{S}_n := E[s_n]$. Note that the PGFs for l_n and s_n in SICTA/F1 can also be derived as in [7], [8], and thus other moments can be readily computed from the derived PGFs, if needed. Notice also that we consider imperfect packet detection in SICTA/F1 over *fading* channels, i.e., we have detection probability $P_D < 1$ and false alarm probability $P_F > 0$. This is different from [8], where SICTA/FS performance is derived only for *AWGN channels*.

1) CRI length: When an idle slot is correctly detected in the presence of fading, SICTA/F1 terminates the CRI. However, if it is falsely regarded as a collision (with probability (w.p.) P_F), then SICTA/FS requires nonexistent "users" to further split, which would surely give rise to an idle subsequent slot. Hence, we have in SICTA/F1 that

$$l_0 = \begin{cases} 1, & \text{w.p. } 1-P_F; \\ 1+l_0, & \text{w.p. } P_F, \end{cases} \tag{14}$$

which in turn yields $\overline{L}_0 = 1/(1-P_F)$.

Over a Rayleigh fading channel, a single packet transmission per slot leads to a success only if it is detected w.p. P_D and successfully decoded w.p. $1-P_e^{(0)}$. Notice that here $P_e^{(0)}$ is the average PER given by (4). If this single packet is not detected w. p. $1-P_D$, the slot is regarded as idle and the CRI ends without success. Otherwise, when it is correctly detected but erroneously decoded w. p. $P_D P_e^{(0)}$, the AP views the slot as a collision slot and enforces further splitting to resolve this "collision". As a result, according to the specification of SICTA/F1, we find

$$l_1 = \begin{cases} 1, & \text{w.p. } 1-P_D P_e^{(0)}; \\ 1+l_1, & \text{w.p. } P_D P_e^{(0)} p; \\ 1+l_0, & \text{w.p. } P_D P_e^{(0)}(1-p). \end{cases} \tag{15}$$

where p is the splitting probability. From (15), we consequently have

$$\begin{aligned} \overline{L}_1 &= 1 - P_D P_e^{(0)} + P_D P_e^{(0)} p(1+\overline{L}_1) \\ &\quad + P_D P_e^{(0)}(1-p)(1+\overline{L}_0) \\ \Rightarrow\ \overline{L}_1 &= \frac{1+\frac{P_D P_e^{(0)}(1-p)}{1-P_F}}{1-P_D P_e^{(0)} p}. \end{aligned} \tag{16}$$

Since we assume that a collision is always correctly detected, it follows that

$$l_n = 1 + l_i, \quad \text{w.p. } B_{n,p}^i, \quad 0 \le i \le n; \quad n \ge 2, \tag{17}$$

where i is the number of users joining the right sub-tree. Using $\overline{L}_0$ and $\overline{L}_1$, we can calculate $\overline{L}_n$ for $n \ge 2$ as [c.f. (17)]

$$\overline{L}_n = \sum_{i=0}^{n} B_{n,p}^i (1+\overline{L}_i) \ \Rightarrow\ \overline{L}_n = \frac{1+\sum_{i=0}^{n-1} B_{n,p}^i \overline{L}_i}{1-p^n}. \tag{18}$$

Summarizing, we have following proposition.

Proposition 2: *The expected values of the conditional CRI length $\overline{L}_n$ in SICTA/F1 can be recursively obtained as*

$$\overline{L}_0 = 1/(1-P_F); \qquad \overline{L}_1 = \frac{1+\frac{P_D P_e^{(0)}(1-p)}{1-P_F}}{1-P_D P_e^{(0)} p}; \tag{19}$$

$$\overline{L}_n = \frac{1+\sum_{i=0}^{n-1} B_{n,p}^i \overline{L}_i}{1-p^n}, \quad \textit{for } n \ge 2; \tag{20}$$

2) Number of Successes: We define the number of successes in a CRI as the number of successfully decoded packets at the end of a CRI. In SICTA/F1, it is clear that $s_0 = 0$; i.e., $\overline{S}_0 = 0$. When there is a single packet transmitted in a CRI, three possibilities arise: 1) the packet is detected and correctly decoded w.p. $P_D(1-P_e^{(0)})$, $s_1 = 1$; 2) the packet is detected but the decoding fails and further splitting gives rise to another single packet transmission w.p. $P_D P_e^{(0)} p$, for which we encounter the same condition again; and 3) the packet is not detected, or, further splitting after the failed decoding gives rise to an idle slot, for which we have $s_1 = 0$. Hence, the specification of SICTA/F1 implies that

$$s_1 = \begin{cases} 1, & \text{w.p. } P_D(1-P_e^{(0)}); \\ s_1, & \text{w.p. } P_D P_e^{(0)} p; \\ 0, & \text{w.p. } 1 - P_D + P_D P_e^{(0)}(1-p); \end{cases} \tag{21}$$

and thus, $\overline{S}_1 = P_D(1-P_e^{(0)}) + P_D P_e^{(0)} p \overline{S}_1$, which yields $\overline{S}_1 = (P_D(1-P_e^{(0)}))/(1-P_D P_e^{(0)} p)$. Note that $\overline{S}_1$ is also the probability that the single packet transmitted in the beginning of a CRI is successfully decoded at the end.

Given $n \geq 2$ initially collided packets and $n-1$ users falling into the right sub-tree, SIC can possibly decode the single packet in the left sub-tree with a certain probability $P_{a,n}$, which is determined by the following lemma.

Lemma 3: *With PER* $P_e^{(j)}$ *after* $j \in [1, n-1]$ *SICs, we have*

$$P_{a,n} = \begin{cases} (1-P_e^{(1)})\frac{B_{2,p}^1}{1-p^2}\overline{S}_1, & n = 2 \\ (1-P_e^{(n-1)})\frac{B_{2,p}^1}{1-p^2}\overline{S}_1 \prod_{j=2}^{n-1} \frac{B_{j,p}^{j-1}}{1-p^j}(1-P_e^{(j-1)}), & n > 2. \end{cases}$$

Proof: For $n = 2$, if binary splitting results in $i = 0$ user into the right-subset, no packet can be successfully decoded; if $i = 1$ user falls into the right-subtree, the other packet of the user in the left sub-tree may be obtained by SIC after the first packet is decoded; if both users fall into the right-subtree, we encounter the same initial collision. Therefore, only when the binary splitting produces a sequence of a slot(s) with 2 collided packets followed by a slot(s) with 1 packet, we may use SIC to decode a extra packet after the first success. This happens with probability $P_{2,1} = B_{2,p}^1 + B_{2,p}^2(B_{2,p}^1 + B_{2,p}^2(B_{2,p}^1 + \cdots)) = B_{2,p}^1(1+p^2+(p^2)^2+\cdots) = \frac{B_{2,p}^1}{1-p^2}$. Besides this specific pattern, we also require successful decoding of the first packet w.p. $\overline{S}_1$ and that of the second packet w.p. $1-P_e^{(1)}$. In this case, we find $P_{a,2} = (1-P_e^{(1)})B_{2,p}^1\overline{S}_1/(1-p^2)$.

For $n > 3$ and $n-1$ users falling into the right sub-tree, successful decoding through SIC again requires that the underlying tree structure follows a specific pattern. Specifically, the right sub-tree with $n-1$ users should consist of a slot(s) with j collided packets followed by a slot(s) with $j-1$ collided packets for every $j \in [2, n-1]$. Notice that between the slot with j collided packets and that with $j-1$ collided packets, slots with the same j collided packets can be present. The probability $P_{j,j-1}$ of this event caused by the binary splitting is then given by

$$\begin{aligned} P_{j,j-1} &= B_{j,p}^{j-1} + B_{j,p}^{j}(B_{j,p}^{j-1} + B_{j,p}^{j}(B_{j,p}^{j-1} + \cdots)) \\ &= B_{j,p}^{j-1}(1 + p^j + (p^j)^2 + \cdots) = \frac{B_{j,p}^{j-1}}{1-p^j}. \end{aligned} \tag{22}$$

Besides the specific pattern required for the right sub-tree, successful decoding in each cancellation step of the SIC occurs with probability $1-P_e^{(j)}$, $j \in [1, n-1]$. Summarizing, we have that for $n > 2$, $P_{a,n} = (1-P_e^{(n-1)})\frac{B_{2,p}^1}{1-p^2}\overline{S}_1 \prod_{j=2}^{n-1} \frac{B_{j,p}^{j-1}}{1-p^j}(1 - P_e^{(j-1)})$. ☐

Using Lemma 3, we find for $n \geq 2$

$$s_n = \begin{cases} s_i, & 0 \leq i \leq n-2, \text{ w.p. } B_{n,p}^i; \\ s_{n-1}+1, & i = n-1, \text{ w.p. } B_{n,p}^{n-1}P_{a,n}; \\ s_{n-1}, & i = n-1, \text{ w.p. } B_{n,p}^{n-1}(1-P_{a,n}). \\ s_n, & i = n, \text{ w.p. } B_{n,p}^n; \end{cases} \tag{23}$$

It thus follows from (23) that for $n \geq 2$

$$\begin{aligned} &\overline{S}_n = B_{n,p}^n\overline{S}_n + B_{n,p}^{n-1}(\overline{S}_{n-1} + P_{a,n}) + \sum_{i=1}^{n-2} B_{n,p}^i\overline{S}_i; \\ &\Rightarrow \ \overline{S}_n = \frac{B_{n,p}^{n-1}P_{a,n} + \sum_{i=1}^{n-1} B_{n,p}^i\overline{S}_i}{1-p^n}. \end{aligned} \tag{24}$$

Proposition 3: *The expected values of the conditional number of successes per CRI* $\overline{S}_n$ *in SICTA/F1 can be recursively obtained as*

$$\overline{S}_0 = 0; \qquad \overline{S}_1 = \frac{P_D(1-P_e^{(0)})}{1-P_D P_e^{(0)} p}; \tag{25}$$

$$\overline{S}_n = \frac{B_{n,p}^{n-1}P_{a,n} + \sum_{i=1}^{n-1} B_{n,p}^i\overline{S}_i}{1-p^n}, \quad \textit{for } n \geq 2; \tag{26}$$

B. BEB-SICTA/F1 and GBEB-SICTA/F1

Since SICTA/F1 largely truncates the original SICTA protocol, it is expected to perform well only when the number of initially collided packets is small. For performance analysis, we employ BEB or GBEB as the channel access algorithm for SICTA/F1. The rules for these combinations are similar as those in (G)BEB-SICTA/FS; see [19] for details. Using the framework in Sec. III-B, we can similarly estimate the MST of BEB-SICTA/F1 with its saturation throughput. First with BEB parameters W, m, p_c and τ defined in Sec. III-B, we can establish that:

Lemma 4: *For a saturation system with* N *users accessing a Rayleigh fading channel, the parameters* p_c *and* τ *in BEB-SICTA/F1 can be found by solving the non-linear equations*

$$\tau = \frac{2(1-2p_c)}{(1-2p_c)(W+1) + p_c W(1-(2p_c)^m)}, \tag{27}$$

$$p_c = 1 - (1-\tau)^{N-1}(P_D(1-P_e^{(0)}) + 1 - P_D). \tag{28}$$

Proof: Eq. (27) can be obtained as in Lemma 2. In SICTA/F1, when a user transmits its packet at the beginning of a CRI and obtains a success (1) feedback after the first slot, it resets its BW size to BW_{min}; otherwise, it doubles its BW size. The success feedback after the first slot of a non-idle CRI occurs only when a single user transmits in the slot and its packet is either detected and successfully decoded, or is undetected by the AP. For a particular user, this happens with probability $(1-\tau)^{N-1}(P_D(1-P_e^{(0)}) + 1 - P_D)$, which leads to eq. (28). ☐

Using τ calculated from (27) and (28), and the CRI statistics $\overline{L}_n$ and $\overline{S}_n$ from Propositions 2 and 3, we can obtain the saturation throughput of BEB-SICTA/F1 over Rayleigh fading

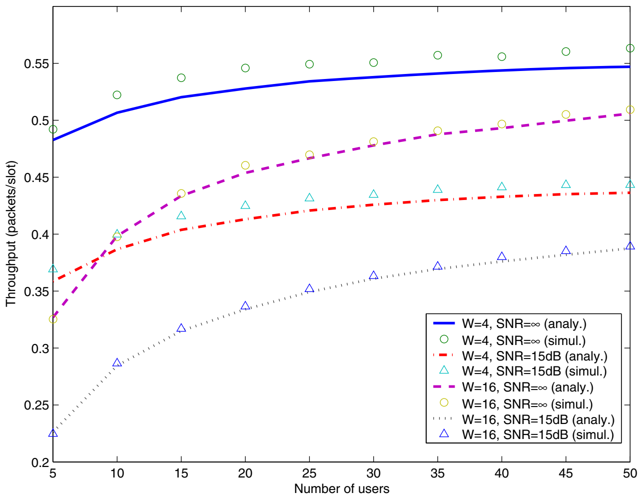


Fig. 4. Comparison between analytical and simulated results for the saturation throughput of BEB-SICTA/F1 over Rayleigh fading channels.

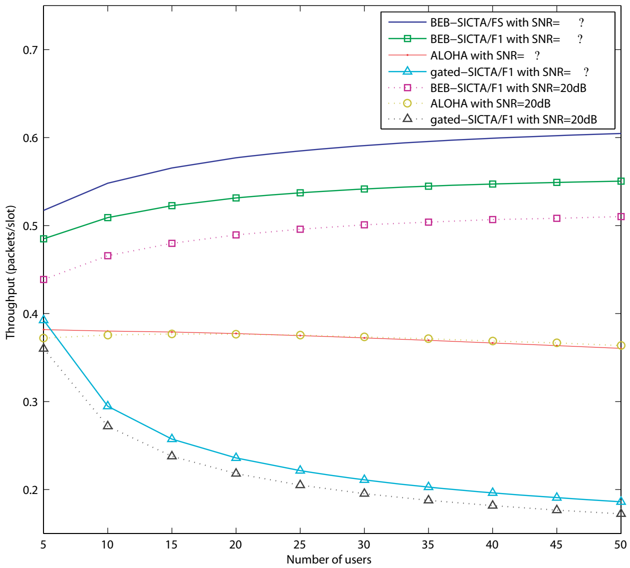


Fig. 5. Saturation throughput comparison among ALOHA (BEB), gated-SICTA/F1 and BEB-SICTA/F1 over Rayleigh fading channels.

channels by Proposition 1. To valid this analysis, we test the BEB-SICTA/F1 protocol on the simulated saturation system for a Rayleigh fading channel with two SNR($:= E_b/N_0$) values : ∞ dB and 15 dB. Fig. 4 compares analytical with simulated results using the minimum BW size $W = 4$ and shows the same comparison when $W = 16$. The analysis is more accurate with a larger W, as in BEB-SICTA/FS. We next compare the saturation throughput of ALOHA (BEB), gated-SICTA/F1 and BEB-SICTA/F1 over Rayleigh fading channels. As depicted in Fig. 5, BEB-SICTA/FS clearly outperforms the others and achieves a 0.55 MST *without* packet loss at infinite SNR. The saturation throughput of BEB-SICTA/FS at infinite SNR is also shown. Although BEB-SICTA/FS yields a higher MST at high SNR, it suffers from the deadlock problem especially at medium or low SNR. By contrast, BEB-SICTA/F1 provides fairly high MST without deadlock across all SNR values. The comparison between GBEB-SICTA/F1 and BEB-SICTA/F1 is similar to that of the corresponding SICTA/FS protocols. By using GBEB as the channel access algorithm for SICTA/F1, GBEB-SICTA/F1 can provide similar saturation throughput across all N and noticeably outperforms BEB-SICTA/F1 when number of users $N \leq 25$, as verified by the simulations; see [19].

V. EXTENSIONS TO IEEE 802.16 WITH MULTIPLE CO SLOTS PER RIE

So far, we developed our (G)BEB-SICTA/FS and (G)BEB-SICTA/F1 protocols for a simple IEEE 802.16 system where only $n_t = 1$ CO slot is scheduled per RIE. For a more realistic setting, this section considers IEEE 802.16 systems with $n_t > 1$, where the feedback messages are now sent per RIE instead of a per slot basis. In this case, depending on the feedback, some slots in an RIE should be reserved for collision resolution and others for random access coordinated by the channel access algorithm. Suppose for instance that in SICTA/FS the feedback for the 5 CO slots in the last RIE is $\{0, 1, e, 1, 1\}$, and the AP schedules 4 CO slots for the current RIE. Then among the latter, the 1st and 2nd slots are reserved for collision resolution, where only the users involved in the 1st and 3rd slots of the last RIE are permitted to transmit accordingly; while the remaining 3rd and 4th slots are used by other users for random access. By taking into account these modifications, the rules of (G)BEB-SICTA/FS and (G)BEB-SICTA/F1 can adhere to the IEEE 802.16 specifications with $n_t > 1$. Next, we evaluate performance of these modifications on a simulated system, where the AP always schedules $n_t = 5$ CO slots per RIE and each user's buffer is fed with a Poisson source having intensity λ packets/slot.

(G)BEB-SICTA/FS over AWGN channels: The simulations are carried out for an AWGN channel at two SNR values: 8 dB and 6 dB. Fig. 6 compares the performance of ALOHA (BEB), BEB-SICTA/FS and GBEB-SICTA/FS when the number of users is $N = 10$; while Fig. 7 depicts throughput comparison when $N = 20$ (the delay performance is similar in both cases). It is clear that GBEB-SICTA/FS achieves over 70% gain in MST relative to ALOHA. Although somewhat inferior to GBEB-SICTA/FS, BEB-SICTA/FS also has much higher MST than ALOHA. Comparing Fig. 6 (a) with Fig. 7, we deduce that GBEB-SICTA/FS yields almost identical MST for the two N values; whereas BEB-SICTA/FS yields smaller MST when $N = 10$ than when $N = 20$. Consequently, the MST gap between GBEB-SICTA/FS and BEB-SICTA/FS increases as N decreases. This observation is corroborated by the results depicted in Fig. 3. From the delay comparisons in Figs. 6 (b), GBEB-SICTA/FS always yields smaller packet delays than ALOHA and BEB-SICTA/FS. Although BEB-SICTA/FS exhibits slightly longer delay than ALOHA at low throughput, it outperforms ALOHA in terms of delay when the throughput is high.

(G)BEB-SICTA/F1 over Rayleigh fading channels: The performance of ALOHA (BEB), BEB-SICTA/F1 and GBEB-SICTA/F1 is evaluated for a Rayleigh fading channel with SNR = 20 dB. Fig. 8 depicts the comparison when the number of users is $N = 20$. GBEB-SICTA/F1 achieves around 50% gain in MST relative to BEB. Although inferior to GBEB-SICTA/F1, BEB-SICTA/F1 has higher MST than BEB. Notice that the MST gap between GBEB-SICTA/F1 and

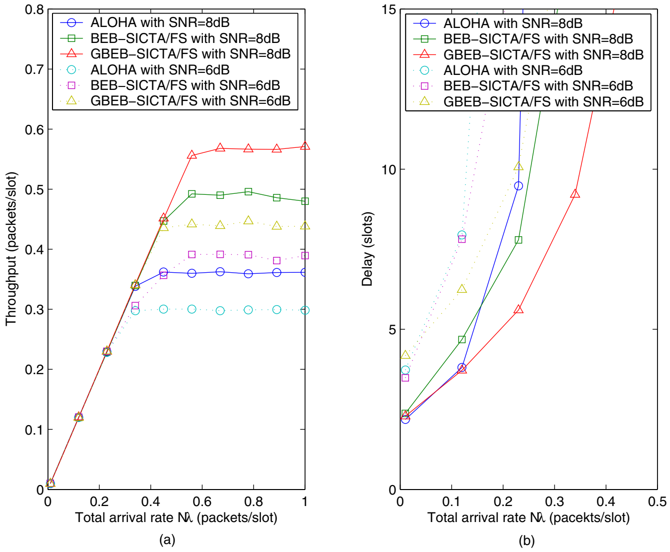


Fig. 6. Performance comparison among ALOHA (BEB), BEB-SICTA/FS and GBEB-SICTA/FS in an IEEE 802.16 system with $N = 10$ users over AWGN channels.

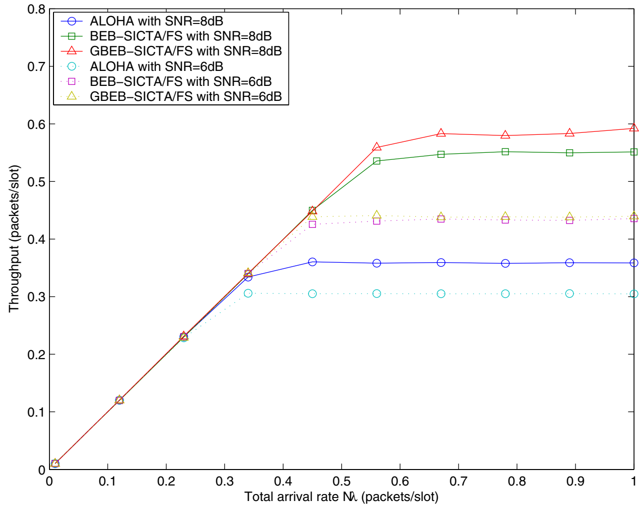


Fig. 7. Performance comparison among ALOHA (BEB), BEB-SICTA/FS and GBEB-SICTA/FS in an IEEE 802.16 system with $N = 20$ users over AWGN channels.

BEB-SICTA/F1 increases as N decreases. Fig. 8 (b) shows that GBEB-SICTA/F1 yields better delay performance than ALOHA, whereas BEB-SICTA/F1 exhibits slightly longer delay than ALOHA.

When SIC is not available at the physical layer, SICTA naturally degrades to MTA, and accordingly our BEB-SICTA/FS and GBEB-SICTA/FS reduce to BEB-MTA/FS and GBEB-MTA/FS. We tested the latter protocols in the simulated IEEE 802.16 system with $N = 20$ users over AWGN channels to assess the performance degradation. As shown in Fig. 9, for SNR=8 dB, the BEB-MTA/FS can achieve almost the same (or a little better) MST as GBEB-MTA/FS and both exhibit about 15% gain in MST relative to ALOHA. However for SNR=6 dB, while GBEB-MTA/FS still enjoys 15% MST gain over ALOHA, the BEB-MTA/FS is no longer superior to ALOHA. We also verify from simulations (which are omitted for conciseness) that GBEB-MTA/FS exhibits a little better delay performance than ALOHA, but BEB-MTA/FS is inferior to ALOHA. The simulation results justify that even without migrating SIC benefits from the physical layer, we may still harness some performance gains when combining BEB or GBEB with MTA/FS. However, these gains are not as significant as those emerging with BEB-SICTA/FS and GBEB-SICTA/FS. The same trend is present for (G)BEB-SICTA/F1 when SIC is not available.

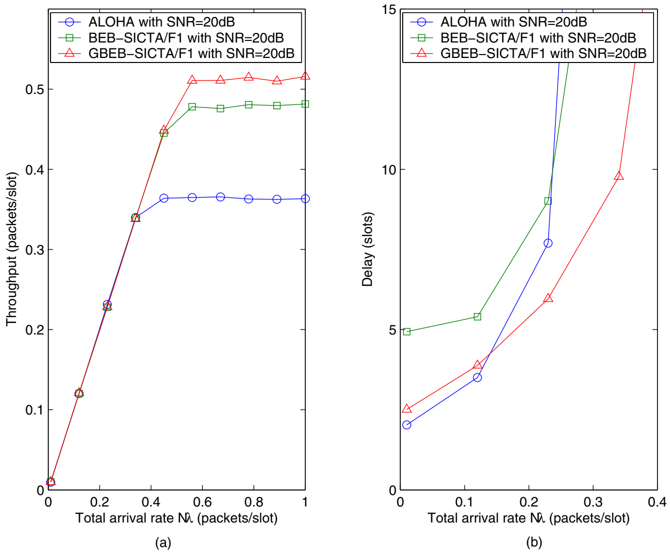


Fig. 8. Performance comparison among ALOHA (BEB), BEB-SICTA/F1 and GBEB-SICTA/F1 in an IEEE 802.16 system with $N = 20$ users over Rayleigh fading channels.

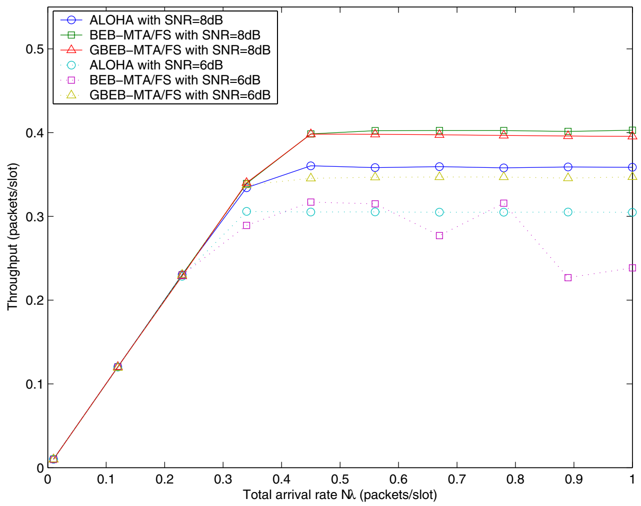


Fig. 9. Performance comparison among ALOHA (BEB), BEB-MTA/FS and GBEB-MTA/FS in an IEEE 802.16 system with $N = 20$ users over AWGN channels.

Some final comments are in order. (G)BEB-SICTA/FS and (G)BEB-SICTA/F1 clearly improves performance of the IEEE 802.16 system well beyond what the currently recommended ALOHA protocol provides. Furthermore, complexity in the novel protocols is affordable because random backoff typically reduces the initial collision size per CRI. As a result, the collision resolution process effected by SICTA/FS or SICTA/F1 is relatively simple. Although (G)BEB-SICTA/FS may provide higher MST, (G)BEB-SICTA/F1 is simpler and more robust by completely avoiding the deadlock in fading channels. While GBEB-type protocols offer a better choice than BEB-type ones from a performance perspective, BEB-SICTA/FS and BEB-SICTA/F1 have advantages in compatibility since they do not need to modify the BEB.

VI. CONCLUSIONS

In this paper, we introduced novel random access protocols for IEEE 802.16 BWA systems. Designed with a cross-layer approach, the novel protocols are robust to errors; they only require limited-sensing; and can attain high MST. Analysis and simulations corroborated that the proposed protocols clearly outperform the ALOHA (BEB) protocol and thereby offer viable alternatives to ALOHA for the random access part of the IEEE 802.16 family of standards.[1]

APPENDIX: PROOF OF LEMMA 2

Upon defining $W_i := 2^i W$, where $i \in [0, m]$ is the so-called "backoff stage," we model the BEB behavior of a user using a state pair $(s(t_c), b(t_c))$, where $b(t_c)$ and $s(t_c)$ denote the stochastic processes representing the BC value and the backoff stage of a given user at t_c. With p_c, the bi-dimensional process $(s(t_c), b(t_c))$ follows a Markov chain depicted in [16, Fig. 4], where the only non null one-step transition probabilities are [16]:

$$\begin{cases} P(i,k|i,k+1) = 1, & k \in [0, W_i - 2], i \in [0, m], \\ P(0,k|i,0) = (1-p_c)/W_0, & k \in [0, W_0 - 1], i \in [0, m], \\ P(i,k|i-1,0) = p_c/W_i, & k \in [0, W_i - 1], i \in [1, m], \\ P(m,k|m,0) = p_c/W_m, & k \in [0, W_m - 1], \end{cases} \tag{29}$$

with $P(i_1, k_1|i_0, k_0) := \Pr(s(t_c+1) = i_1, b(t_c+1) = k_1 | s(t_c) = i_0, b(t_c) = k_0)$. If $b_{i,k} := \lim_{t_c \to \infty} P(s(t_c) = i, b(t_c) = k)$, $\ i \in [0, m]$, $k \in [0, W_i - 1]$, then $b_{i,k}$ can be found in closed-form as [16]

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot \begin{cases} (1-p_c)\sum_{j=0}^{m} b_{j,0}, & i = 0, \\ p_c b_{i-1,0}, & 0 < i < m, \\ p_c(b_{m-1,0} + b_{m,0}), & i = m; \end{cases} \tag{30}$$

$$\Rightarrow\ b_{0,0} = \frac{2(1-2p_c)(1-p_c)}{(1-2p_c)(W+1) + p_c W(1-(2p_c)^m)}. \tag{31}$$

Since a user transmits when its BC value reaches zero, the probability τ is given by

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{b_{0,0}}{1-p_c} = \frac{2(1-2p_c)}{(1-2p_c)(W+1) + p_c W(1-(2p_c)^m)}.$$

On the other hand, collision occurs when more than two users transmit at the same time. Different to [16], we also regard the packet transmission failures caused by AWGN as "collisions" (since the AP cannot distinguish these failures from those caused by real collisions), and take the PER associated with these "collisions" into account. Then using τ and PER $P_e^{(0)}$, we can write $p_c = 1 - (1-\tau)^{N-1}(1 - P_e^{(0)})$ which completes the proof.

REFERENCES

[1] J. I. Capetanakis, "Tree algorithm for packet broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 505-515, Sept. 1979.

[2] B. S. Tsybakov and V. A. Mikhailov, "Tree algorithm for packet broadcast channels," *Problemy Peredachi Informatsii*, vol. 14, no. 4, pp. 32-59, Oct.-Dec. 1978.

[3] M. L. Molle and G. C. Polyzos, "Conflict resolution algorithms and their performance analysis," Department of Computer Science and Engineering, University of California, San Diego, CA, Tech. Rep. CS93-300, July 1993.

[4] B. S. Tsybakov and V. A. Mikhailov, "Random multiple packet access: part-and-try algorithm," *Problemy Peredachi Informatsii*, vol. 16, no. 4, pp. 65-79, Oct.-Dec. 1980.

[5] B. S. Tsybakov and N. Likhanov, "Upper bound on the capacity of a random access system," *Problemy Peredachi Informatsii*, vol. 23, no. 3, pp. 64-78, July-Sept. 1987.

[6] M. K. Tsatsanis, R. Zhang, and S. Banerjee, "Network assisted diversity for random access wireless systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 702-711, March 2000.

[7] Y. Yu and G. B. Giannakis, "SICTA: a 0.693 contention tree algorithm using successive interference cancellation," *Proc. INFOCOM*, vol. 3, pp. 13-17, March 2005.

[8] X. Wang, Y. Yu, and G. B. Giannakis, "A robust high-throughput tree algorithm using successive interference cancellation," *Proc. GLOBECOM*, Nov. 2005.

[9] J. G. Proakis, *Digital Communications, 3rd ed.* New York: McGraw-Hill, 1995.

[10] J. L. Massey, *Collision Resolution Algorithm and Random Access Communications*. Multiuser Communication Systems 265, ed. G. Longo, CISM Course and Lecture Notes, Spriger, New York, 1981.

[11] D. Bertsekas and R. Gallager, *Data Networks, 2nd ed.* Upper Saddle River, NJ: Prentice Hall, 1992.

[12] IEEE 802.16 WG, *Air interface for fixed broadband wireless access systems*, IEEE Std. 802.16, April. 2002.

[13] R. Garces and J. Garcia-Luna-Aceves, "Collision avoidance and resolution multiple access: first-success protocols," *Proc. Intl. Com. Conf.*, vol. 2, pp. 699-703, June 1997.

[14] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746-1755, Sept. 2004.

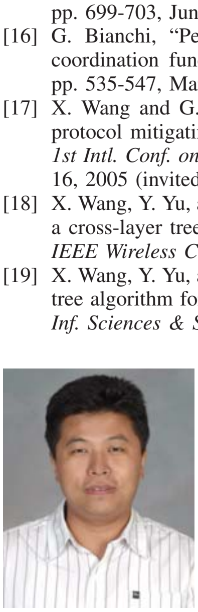
[15] R. Garces and J. Garcia-Luna-Aceves, "Collision avoidance and resolution multiple access: first-success protocols," *Proc. IEEE ICC*, vol. 2, pp. 699-703, June 1997.

[16] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535-547, Mar. 2000.

[17] X. Wang and G. B. Giannakis, "CSMA/CCA: a modified CSMA/CA protocol mitigating the fairness problem for IEEE 802.11 DCF," *Proc. 1st Intl. Conf. on Multimedia Services Access Net.*, pp. 88-95, June 12-16, 2005 (invited).

[18] X. Wang, Y. Yu, and G. B. Giannakis, "Combining random backoff with a cross-layer tree algorithm for random access in IEEE 802.16," *Proc. IEEE Wireless Communications & Networking Conf.*, April, 2006.

[19] X. Wang, Y. Yu, and G. B. Giannakis, "A deadlock-free high-throughput tree algorithm for random access over fading channels," *Proc. Conf. on Inf. Sciences & Systems*, March 22-24, 2006.



Xin Wang (S'03-M'04) received his B.Sc. in 1997 and M.Sc. in 2000 from Fudan University, Shanghai, China, and his Ph.D. degree in 2004, from Auburn University all in Electrical Engineering.

From September 2004 to August 2006, he was a postdoctoral research associate with Department of Electrical and Computer Engineering, University of Minnesota. Since September 2006, he has been an assistant professor at the Department of Electrical Engineering, Florida Atlantic University. His research interests include medium access control, cross-layer design, resource allocation and signal processing for communication networks.

---

[1] The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.

Yingqun Yu received his B.S. degree from Nanjing University, Nanjing, China, in 1997, M.S. degree from Tsinghua University, Beijing, China, in 2000, and PhD degree from University of Minnesota, Twin Cities, Minnesota, in 2007, all in electrical engineering.

Since May 2007, he has been with Qualcomm Inc., San Diego, CA. His research interests include information theory, wireless communications and networking, and signal processing for communications.

Georgios B. Giannakis (S'84-M'86-SM'91-F'97) received his Diploma in Electrical Engineering from the National Technical University of Athens, Greece, 1981. From September 1982 to July 1986 he was with the University of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engineering, 1986. After lecturing for one year at USC, he joined the University of Virginia in 1987, where he became a professor of Electrical Engineering in 1997. Since 1999 he has been a professor with the Department of Electrical and Computer Engineering at the University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications.

His general interests span the areas of communications, networking, and statistical signal processing - subjects on which he has published more than 250 journal papers, 450 conference papers and four books. Current research focuses on diversity techniques for fading channels, complex-field and space-time coding, multicarrier, ultra-wide band and cognitive radios, cross-layer designs, ad hoc and wireless sensor networks.

G. B. Giannakis is the (co-) recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, 2004). He also received Technical Achievement Awards from the SP Society in 2000 and from EURASIP in 2005. He served as Editor in Chief for the *IEEE SP Letters*, as Associate Editor for the *IEEE Transactions on Signal Processing* and the *IEEE SP Letters*, as secretary of the SP Conference Board, as member of the SP Publications Board, as member and vice-chair of the Statistical Signal and Array Processing Technical Committee, as chair of the SP for Communications Technical Committee and as a member of the IEEE Fellows Election Committee. He has also served as a member of the IEEE-SP Society's Board of Governors, the Editorial Board for the *Proceedings of the IEEE* and the steering committee of the *IEEE Transactions on Wireless Communications*.