Hexanary-Feedback Contention Access With PDF-Based Multiuser Estimation for Wireless Access Networks

Maria C. Yuang, *Senior Member, IEEE*, Bird C. Lo, *Member, IEEE*, and Ju-Ya Chen, *Member, IEEE*

***Abstract*—Most existing contention access schemes are inherently unstable resulting in exponentially deteriorating throughput under increased traffic loads. In this paper, we propose a *wide-sense stable* (WSS) efficient *hexanary-feedback contention access* (HFCA) scheme, capable of providing signaling traffic high performance while retaining maximal throughput for wireless access networks. HFCA performs incremental contention resolution, managing a small subset of users at a time via a two-phase process. In the first phase, a group of users is probabilistically admitted, with a negligible probability of the group size greater than five. In the second phase, all users in the group are efficiently resolved. The two-phase process is augmented with hexanary feedback control facilitated by a probability density function (pdf)-based multiuser estimator (PMER) implemented at the physical layer. Basically, PMER measures the exact number of transmitting users (zero to five) in a contention slot by matching the envelope–phase pdf's histograms of received signals to a preconstructed pdf's library. To formally justify the performance of HFCA, we present throughput and stability analysis in which HFCA is shown WSS and the strict-sense stability condition is derived. Finally, analytic and simulation results delineate that, HFCA is highly robust against estimation discrepancy. Significantly, HFCA achieves high performance with respect to maximum stable and saturated throughputs, access delay, and blocking probability.**

***Index Terms*—Contention access, envelope–phase probability density function (pdf), feedback control, maximum stable throughput, saturated throughput, wireless access networks.**

I. Introduction

WIRELESS access networks are expected to support multiple services [1] with a wide range of service rates and different quality of service (QoS) requirements. Expected supported services include constant bit rate (CBR), variable bit rate (VBR), available bit rate (ABR), and in-band signaling traffic for making bandwidth reservation for above traffic. Examples of QoS requirements for CBR/VBR, ABR, and signaling traffic are bounded jitter–delay, minimum cell rate, and maximum blocking probability, respectively. It has been shown that the former three types of guaranteed (or semiguaranteed) traffic could be efficiently governed by reservation access [2]. The signaling traffic, on the other hand, is most suitably directed by contention access [2]–[4]. It is well known that contention access incurs inevitable collisions, resulting in throughput deterioration and dissatisfaction of QoS requirements under increased traffic loads. A key challenge pertaining to such wireless access networks has been the design of contention access satisfying access efficiency and QoS guarantees [3].

Existing time-division multiple access (TDMA) [5], [6] methods can be categorized as either frequency-division duplex (FDD) [5], or time-division duplex (TDD) [6]. In FDD-based systems, uplink and downlink traffic are carried by two distinct carrier frequencies. In TDD-based systems, while using one common carrier frequency, uplink and downlink traffic are carried at different time intervals. In this paper, we focus on the design of a TDMA FDD-based contention access scheme. Among existing schemes, a set of methods [7], [8] adopts the use of minislots in an effort to maintain system stability for traffic loads up to channel capacity. However, due to high cost for maintaining global minislot synchronization, we disregard the minislot-based approach in our work.

Among prevailing schemes, splitting-based collision-resolution algorithms [9]–[11] have been consider promising. The basic idea behind the design is to speed up the resolution process by probabilistically [9] or time [10], [11], splitting contenders into transmitting and nontransmitting sets based on various types of feedback that is made available to users. There are three types of feedback—binary feedback, ternary feedback [9], [11], and multiplicity feedback [10]. In binary feedback, users are informed of a binary result—success or collision in a contention slot. In ternary feedback, users are further informed of three outcomes—idle, success, or collision. Finally, multiplicity feedback allows users to have full knowledge of the exact number of users involved in a collision. Compared to binary–ternary feedback, multiplicity feedback enables greatly improved access efficiency, however, at the expense of unmanageable implementation complexity. The goal of the paper is to design an efficient tractable *hexanary feedback-based* contention access scheme, in which hexanary feedback can simply be facilitated in hardware.

Exploiting such feedback control, proposed algorithms [9]–[12] exhibit different merits with respect to four performance metrics: stability, saturated throughput, maximum stable throughput, and implementation complexity, as summarized in Table I. Notice that our intention is not to provide a thorough

Manuscript received January 27, 2002; revised July 10, 2002 and September 21, 2002; accepted October 1, 2002. The editor coordinating the review of this paper and approving it for publication is X. Wang. This work was supported in part by the MOE Program of Excellence Research, Taiwan, R.O.C., under Contract 89-E-FA04-1-4, and in part by the Institute for Information Industry (III), MOEA, Taiwan, R.O.C., under Contract 92-0454.

M. C. Yuang and B. C. Lo are with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu 30050, Taiwan, R.O.C.

J.-Y. Chen is with the Institute of Communication Engineering, National Sun Yat-Sen University, Kaohsiung 804, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TWC.2003.819023

TABLE I
EVALUATION OF FEEDBACK-BASED COLLISION-RESOLUTION ALGORITHMS

Collision-Resolution Algorithm	Feedback Type	Stability	Maximum Stable Throughput	Saturated Throughput	Implementation Complexity
Binary Tree [11]	Ternary	Unstable	0.487	0	Medium
QMAC [12]	Binary	WSS	0.507	0.368	Low
Georgiadis [10]	Multiplicity	Unstable	0.532	0	High
Paris [9]	Ternary	WSS	0.487	0.368	High
HFCA	Hexanary	WSS	0.605	0.522	Medium

survey of existing algorithms. Instead, we aim to introduce salient performance terms and, in turn, highlight the impact of different designs on these performance metrics. First of all, a medium access control (MAC) scheme is classified as *wide-sense stable* (WSS) [3], [9], [12] if the network retains goodput even when the system is saturated. The positive throughput under the saturated condition is called the saturated throughput. Next, both WSS and unstable [10], [11] schemes experience stable behavior when the traffic arrival rate is below a threshold. The maximum achievable throughput under the stable condition is referred to as the maximum stable throughput. Furthermore, a scheme is classified as *strict-sense stable* (SSS) if not only it is WSS but each user's queue is retained stable. Finally, implementation complexity is often incurred by operations requiring full sensing of feedback [11], computation of transmission–retransmission probability functions [9], and hardware energy detectors [10].

In this paper, we propose a WSS efficient *hexanary-feedback contention access (HFCA)* scheme, capable of providing signaling traffic high performance while retaining maximal throughput for wireless access networks. HFCA performs incremental contention resolution, managing a small subset of users at a time via a two-phase process. In the first phase, a group of users is probabilistically admitted, with a negligible probability of the group size greater than five. In the second phase, all users in the group are efficiently resolved. The two-phase process is augmented with hexanary feedback control facilitated by a probability density function (pdf)-based multiuser estimator (PMER) implemented at the physical layer.

By matching the envelope–phase pdf's histograms of received signals to a library of preconstructed pdfs, PMER measures the exact number of users (zero to five) involved in a contention slot. These six outcomes comprise the hexanary feedback that is broadcast to users during the real-time operation of the two-phase process. To formally justify the performance of HFCA, we present throughput and stability analyses. In the analyses, HFCA is shown WSS and the strict-sense stability condition is derived. Finally, analytic and simulation results delineate that HFCA is highly robust against estimation discrepancy. Significantly, HFCA achieves high performance with respect to maximum stable and saturated throughputs, access delay, and blocking probability.

The remainder of this paper is organized as follows. In Section II, we describe the network and system architectures. In Section III, we introduce the design of PMER and demonstrate its estimation performance. In Section IV, we detail the two-phase HFCA scheme, followed by the throughput and stability analyses. Analytic and simulation results are also demonstrated in Section IV. In Section V, we present the robustness assessment for HFCA. Finally, concluding remarks are given in Section VI.

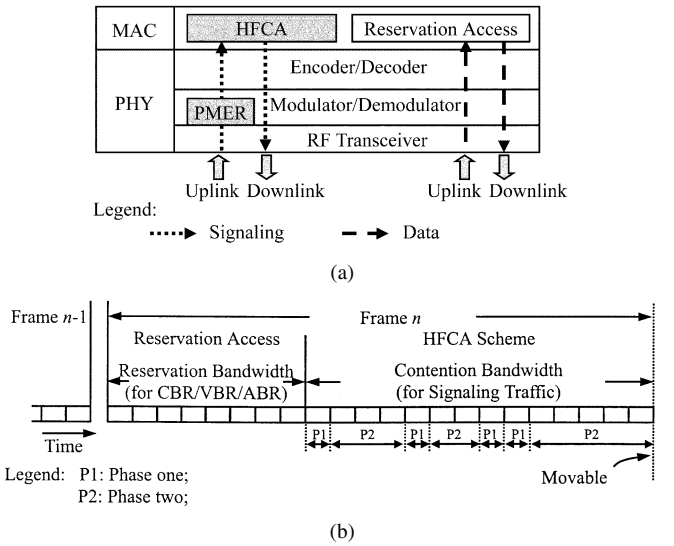


Fig. 1. System architecture and frame structure. (a) System architecture of the base station. (b) Frame structure.

II. NETWORK AND SYSTEM ARCHITECTURES

The wireless network architecture is the classical cell with a base station (BS) serving a finite set of mobile terminals (MTs) by means of a shared radio medium. On the basis of FDD, the medium bandwidth is divided into uplink and downlink channels. The uplink channel transfers information from MTs to the BS according to a TDM-based MAC protocol incorporating our newly proposed HFCA scheme. That is, time on the uplink channel is divided into a contiguous sequence of variable-size TDMA frames each of which is further subdivided into a fixed number of slots. Transmissions from all MTs are assumed independent. The downlink channel typically broadcasts information and acknowledges previous transmissions made on the uplink channel. Its operation is beyond the scope of this paper. Notice that due to FDD and small propagation delay in a local environment, immediate feedback and acknowledgment from the BS via downlink channel can be made available to all MTs prior to the beginning of the subsequent slot of the uplink channel.

The system architecture of the BS consists of two layers—physical (PHY) and MAC layers, as shown in Fig. 1(a). There are three modules (encoder–decoder, modulator–demodulator, and radio-frequency (RF) transceiver) at the PHY layer, and two access schemes (HFCA and reservation access) at the MAC layer. At the PHY layer, for forward downlink traffic, the transmitted digital data is first encoded by the channel encoder, which protects the transmitted data from channel distortion. The encoded signal is then passed through the modulator and, in turn, emitted into the air via the RF transceiver. For reverse uplink traffic, the received signal is down-converted to baseband by the RF transceiver module. In the case of signaling traffic, the baseband signal is delivered into PMER for the estimation of

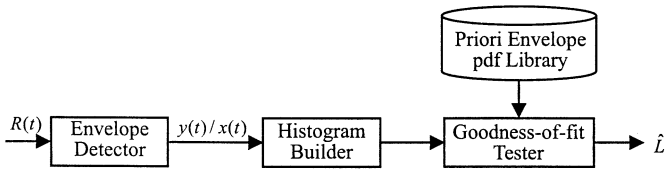


Fig. 2. Design of PMER.

colliding user population size before being demodulated. Otherwise, the signal is directly demodulated and decoded to recover the transmitted data.

At the MAC layer, reservation access manages the access for CBR/VBR/ABR traffic, and the HFCA scheme governs that for signaling traffic, using dynamically allocated reservation and contention bandwidth (in units of slots), respectively, as shown in Fig. 1(b). The operation of reservation access is beyond the scope of the paper. An MT wishing to establish a CBR, VBR, or ABR connection has to make a signaling request in a slot within contention bandwidth based on the HFCA scheme described in Section IV. In the sequel, we first present the design and performance of PMER.

III. PMER

A. Design Concept

The basic design rests on the following fact. Under a given reasonable signal-to-noise ratio (SNR), for received signals contributed by different but small numbers of users, their envelope–phase pdfs are significantly distinctive. Specifically, as will be shown later, if there are less than six users, pdfs of multiuser signals are distinguishable. Otherwise, multiuser signals are Gaussian distributed. Accordingly, one can construct in advance *a priori* pdf library from theoretical derivations, simulation results, and/or field measurements under various SNRs and numbers (one to five) of concurrent users. In our work, we consider two channel models—additive white Gaussian noise (AWGN) and narrowband multipath fading. For the AWGN channel, we employ the envelope pdf on the basis of a theoretically derived pdf library. For the fading channel, we adopt the phase pdf using a simulation-result-based pdf library for simplicity purpose. For other practical channel environments, pdf libraries can be constructed by field measurements.

In the sequel, we describe the design of PMER (see Fig. 2) under the AWGN case. During the on-line operation at the end of each slot, taking the received signal $R(t)$ as the input, the envelope detector first generates the associated envelope $y(t)$ and normalized envelope $x(t)$. The pdf of the normalized envelope is next quantized via a histogram builder, constructing the corresponding normalized envelope histogram. The goodness-of-fit tester then conducts tests of goodness-of-fit for the resulted histogram against the priori envelope library previously constructed. Candidates of test methods include least square (LS), weighted LS, maximum likelihood, and maximum *a posteriori* (MAP) [13]. The output of the tester is the estimated number of concurrent users $\hat{L}$. Notice that the $\hat{L}$ value, ranging from one to five, together with "idle," corresponds to six different outcomes of feedback.

B. Statistics of Received Multiuser Signal

In this subsection, we derive the normalized envelope pdf and phase pdf of the received multiuser signal under the AWGN and narrowband multipath fading channels, respectively.

Case 1: AWGN Channel: During the on-line operation at the end of each slot access, a phase-modulated signal ($R(t)$) received by the BS, which is contributed by L users, can be given as

$$R(t) = \sum_{i=1}^{L} S_i e^{j[2\pi f_0 t + \phi_i(t)]} + n(t) e^{j2\pi f_0 t} \tag{1}$$

where S_i is the square root of the ith MT's power at the receiver, f_0 is the carrier frequency, $\phi_i(t)$ is the modulated waveform from the ith MT, and $n(t)$ represents the Gaussian noise with zero mean and variance $2\sigma_n^2$. For the ease of illustration, we assume that power control is exerted at each MT, leading to $S_i = S$ for all is. Denote $y(t)$ the envelope of $R(t)$. Let $v(t)$ be the real part of $R(t)$. First, we obtain the characteristic function of $v(t)$ as

$$f_v(t) = \prod_{i=1}^{L} J_0(S_i t) e^{-(\sigma_n^2 t^2/2)} = {J_0}^L(St) e^{-(\sigma_n^2 t^2/2)} \tag{2}$$

where $J_0(\cdot)$ is the Bessel function of the first kind of order zero. The pdf of $y(t)$ can then be derived [14] as

$$\begin{aligned} p_y(y) &= \int_0^\infty yt J_0(yt) f_v(t) dt \\ &= \int_0^\infty yt J_0(yt) {J_0}^L(St) e^{-(\sigma_n^2 t^2/2)} dt, y > 0. \end{aligned} \tag{3}$$

To normalize the received signal power according to

$$x^2(t) = \frac{y^2(t)}{\text{mean received power}} = \frac{y^2(t)}{L \cdot S^2 + 2\sigma_n^2} \tag{4}$$

the corresponding pdf of the normalized envelope $x(t)$ can be derived from (3) as

$$\begin{aligned} &p_x(x) \\ &= (L \cdot \Lambda + 1) x \int_0^\infty t e^{-t^2/4} J_0\left(xt\sqrt{L \cdot \Lambda + 1}\right) J_0^L\left(t\sqrt{\Lambda}\right) dt \end{aligned} \tag{5}$$

where $\Lambda = S^2/2\sigma_n^2 = \text{SNR}$. We clearly notice from (5) that $p_x(x)$ is a function of L and SNR. Thus, *a priori* envelope pdf library can be offline-constructed based on (5). Partial results of the library for different numbers of users ($L = 1$ to 5) under $\text{SNR} = 10$ and 20 dB, are depicted in Fig. 3(a) and (b), respectively. Clearly, the stronger the SNR, the more distinctive the pdfs are.

Case 2: Narrowband Multipath Fading Channel: When a received signal undergoes narrowband multipath fading, the signal strength S_i is distributed as a Rician or Rayleigh function. The normalized envelope distribution is also a Rician or Rayleigh function. Hence, the concurrent user number cannot be identified from the normalized envelope distribution. However, we discover that the phase pdf can be used instead in this case.

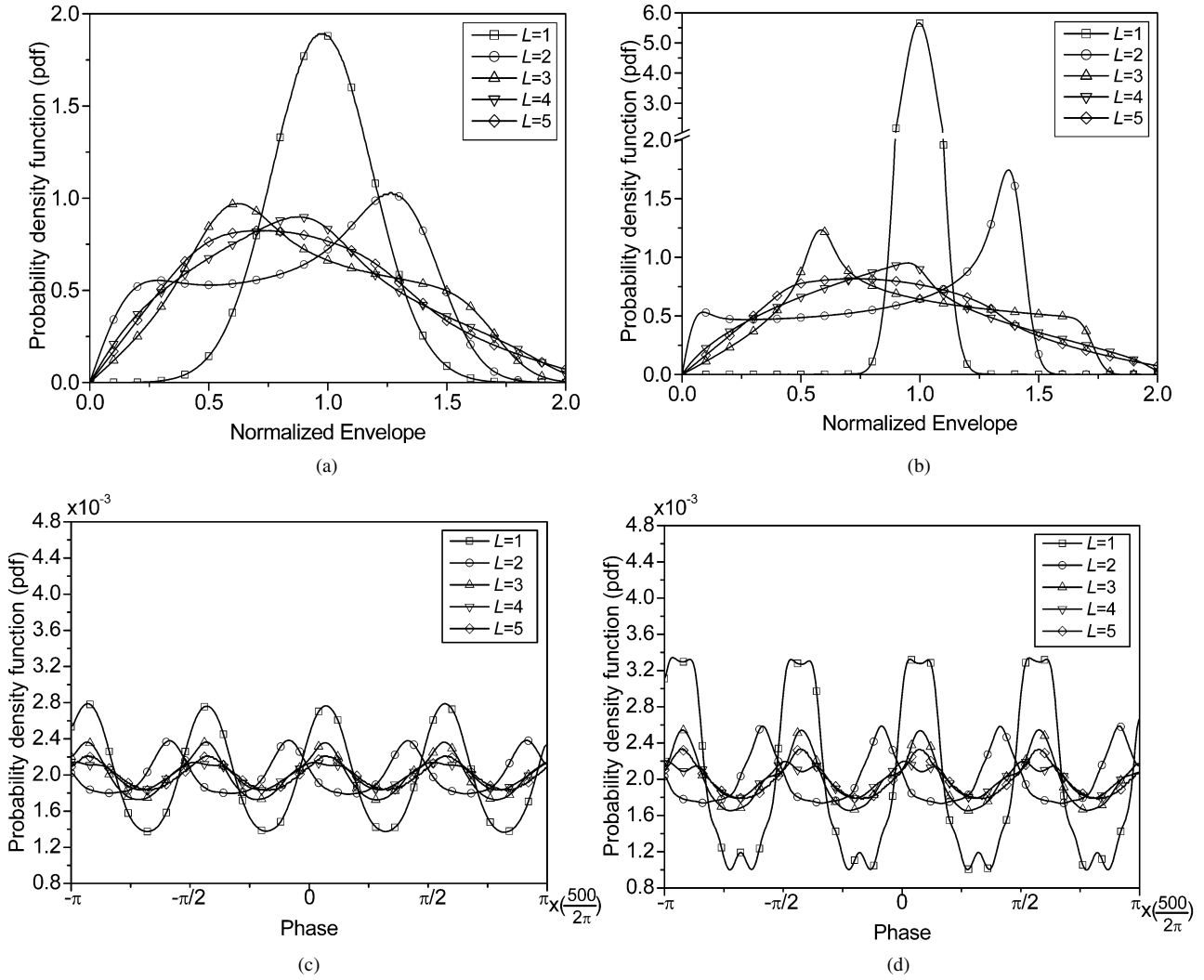


Fig. 3. Envelope–phase pdfs for different Ls. (a) Normalized envelope pdf under $\mathrm{SNR} = 10$ dB. (b) Normalized envelope pdf under $\mathrm{SNR} = 20$ dB. (c) Phase pdf under $\mathrm{SNR} = 10$ dB. (d) Phase pdf under $\mathrm{SNR} = 20$ dB.

First of all, with S_i and ϕ_i normalized by the power and phase of a designated MT, say S_1 and ϕ_i, i.e., $S_i' = S_i/S_1$ and $\phi_i' = \phi_i - \phi_1$, (1) can be reformulated as

$$R(t) = S_1 e^{j[2\pi f_0 t + \phi_1(t)]} \left\{ 1 + \sum_{i=2}^{L} S_i' e^{j\phi_i'(t)} \right\} + n(t) e^{j2\pi f_0 t}. \tag{6}$$

Using [15, eq. (13)], the phase pdf of the received signal $R(t)$ can be derived as (7), shown at the bottom of the page, where θ is the phase with the range between $-\pi$ and π, and I_0 is the zeroth order modified Bessel function of the first kind. With S_i^2 replaced by $\Lambda_i \cdot 2\sigma_n^2$, $p(\theta|S_i)$ in (7) is transformed into a function of SNR (Λ_i), namely (8), shown at the bottom of the page. Finally, by unconditioning $p(\theta|\Lambda_i)$ in (8), we have

$$p(\theta) = \int_0^\infty \cdots \int_0^\infty p(\theta|\Lambda_i) p(\Lambda_1) \ldots p(\Lambda_L) d\Lambda_1 \ldots d\Lambda_L, \tag{9}$$

$$p_\theta(\theta|S_i) = \frac{1}{2\pi} \int_0^\infty \exp\left[-\left(\frac{r^2}{2} + \sum_{i=1}^{L} \frac{S_i^2}{2\sigma_n^2} - \frac{S_1 r \cos\theta}{\sigma_n} \right) \right] \prod_{i=2}^{L} I_0 \left(\frac{S_i}{\sigma_n} \sqrt{r^2 + \frac{S_1^2}{\sigma_n^2} - \frac{2S_1 r}{\sigma_n} \cos\theta} \right) r dr \tag{7}$$

$$p_\theta(\theta|\Lambda_i) = \frac{1}{2\pi} \int_0^\infty \exp\left[-\left(\frac{r^2}{2} + \sum_{i=1}^{L} \Lambda_i - \sqrt{2\Lambda_1} r \cos\theta \right) \right] \prod_{i=2}^{L} I_0 \left(\sqrt{2\Lambda_i r^2 + 4\Lambda_1 \Lambda_i - 4\Lambda_i \sqrt{2\Lambda_1} r \cos\theta} \right) r dr \tag{8}$$

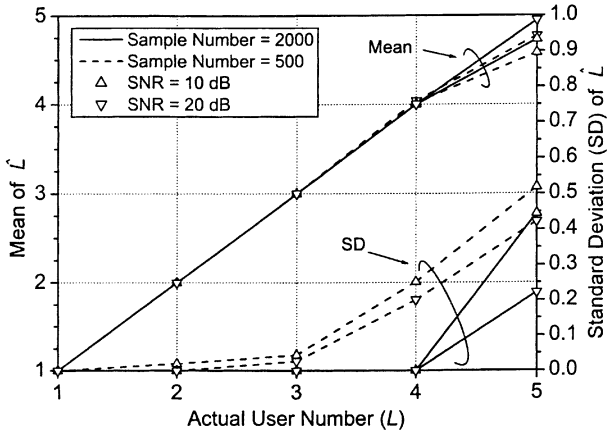


Fig. 4. Mean and standard deviation of the estimated user number ($\hat{L}$).

where $p(\Lambda_i)$ is the pdf of the received signal's SNR. Notice that $p(\theta)$ is a function of L and SNR. Therefore, *a priori* phase pdf library can be similarly constructed for the estimation of $\hat{L}$. Since the closed form of the phase pdf is analytically intractable, we generate the phase pdf via simulation. In the simulation, we employed a 0.3-GMSK modulated signal passed through a narrowband Rayleigh fading channel with Doppler frequency 40 Hz. Simulation results are depicted in Fig. 3(c) and (d). We clearly observe that the pdfs corresponding to different number of users (L) are distinctive.

C. Simulation Results

We carried out an experiment to examine estimation precision via simulation. For simplicity, we only consider the AWGN channel environment in the simulation. During the simulation, 16 uniform partitions were adopted for the quantization within the histogram builder. The simplest LS method was exerted for goodness-of-fit tests. Simulation was operated for a total of 500 slots in length. First of all, based on fixed numbers of samples (500 and 2000) collected at the end of each slot, we interrogated the mean and standard deviation of the estimated user number $\hat{L}$, under $\mathrm{SNR} = 10$ and 20 dB. Simulation results are plotted in Fig. 4. We observe from the figure that the mean estimated $\hat{L}$ values (one to five) using 2000 samples completely agree with the actual L values under $\mathrm{SNR} = 20$ dB. The standard deviation under the worst case ($L = 5$) is as low as 0.222. As for the results based on a smaller sample number, i.e., 500 samples, estimation is profoundly precise particularly under $L \leq 4$.

We carried out another experiment to investigate normalized estimation precision of five different $\hat{L}$ values as a function of the sample number. The normalized estimation precision is defined as the ratio of the number of slots resulting in correct estimation to the total number of slots experimented, which is 500 slots in this case. Results are displayed in Fig. 5. For $\hat{L} = 5$ and under $\mathrm{SNR} = 20$ dB, for example, we discover that PMER achieves precision of 77.8%, 93.95%, and 99.14%, using 500, 2000, and 5000 samples, respectively. We unsurprisingly perceive that estimation precision increases with the sample number. Higher user numbers require more samples to achieve the same grade of precision.

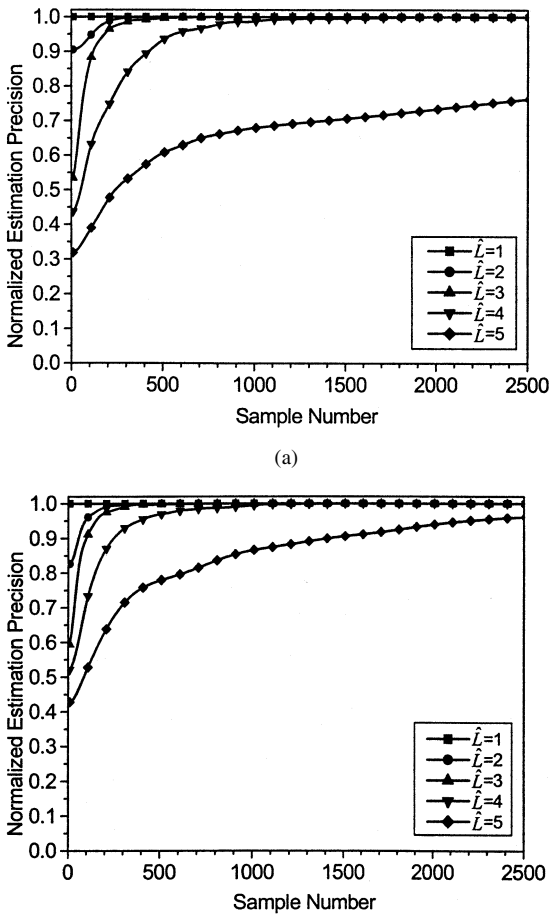


Fig. 5. Normalized estimation precision under different sample numbers. (a) $\mathrm{SNR} = 10$ dB (b) $\mathrm{SNR} = 20$ dB.

Finally, we are at the stage to determine the mean estimation precision of PMER with the occurring probability of each user number taken into account. This can be computed by summing the product of normalized estimation precision of each $\hat{L}$ and the corresponding occurring probability. In the computation, we adopt the framing structure specified in a global sytem for mobile communication (GSM). The system uses 148 transmitted data bits in a burst and a sampling rate of eight samples per bit, resulting in a total of $148 \times 8 = 1184$ samples. With these samples, baseed on results in Fig. 5, PMER exhibits precision of 100% ($\hat{L} = 0$), 100% ($\hat{L} = 1$), 100% ($\hat{L} = 2$), 100% ($\hat{L} = 3$), 99.34% ($\hat{L} = 4$), and 69% ($\hat{L} = 5$) under $\mathrm{SNR} = 10$ dB; and 100% ($\hat{L} = 0$), 100% ($\hat{L} = 1$), 100% ($\hat{L} = 2$), 100% ($\hat{L} = 3$), 99.78% ($\hat{L} = 4$), and 88.04% ($\hat{L} = 5$) under $\mathrm{SNR} = 20$ dB. We next acquire probabilities $p(L = k)$, $k \leq 5$. According to (27) (given later), consider $\hat{N} = 100$, $\delta = 0$, we obtain $p(L = k) = 0.203, 0.327, 0.259, 0.136, 0.053, 0.016$, for $k = 0$ to 5, respectively. After computing, the mean estimation precisions under $\mathrm{SNR} = 10$ and 20 dB, are given as 98.9% and 99.27%, respectively.

IV. HFCA

The basic idea behind the design is to overcome system instability by probabilisticly reducing the contention size to less

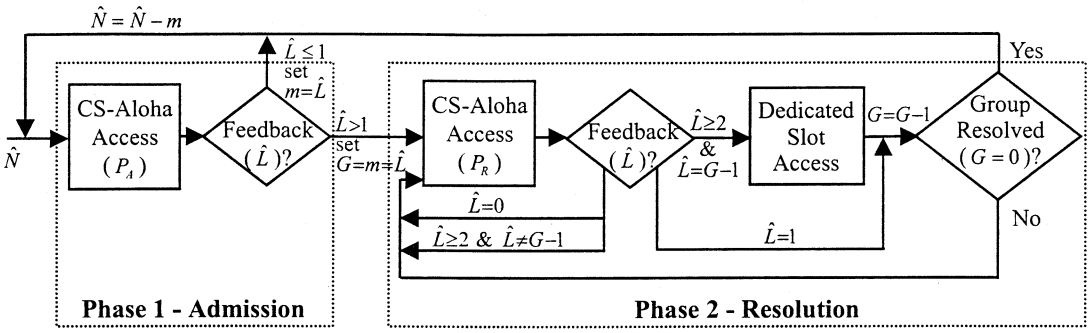


Fig. 6. Two-phase algorithm of HFCA.

than six, followed by efficient collision resolution with the aid of hardware-based hexanary feedback. The hexanary feedback informs six possible outcomes—idle, success, and two- to five-user collisions. In the sequel, we initially describe its operation in detail. We then present the throughput and stability analyses. We demonstrate the performance with respect to throughput, access delay, and blocking probability, via analytic and simulation results. Finally, bringing PMER into play, we assess the robustness of the entire system.

A. Basic Operation

The operation of HFCA within a frame [see Fig. 1(b)] consists of repeated executions of two phases—admission and resolution phases. While a group of MTs is randomly selected in the admission phase, the selected group is resolved in the resolution phase. The two-phase process repeats until either the maximum blocking probability is satisfied (before all users are resolved) or contention bandwidth consumption exceeds the maximum frame boundary [12]. It is worth noting that new users that become active during the current frame's contention period are inhibited from transmitting while the contention process is in progress. The transmissions have to be postponed until the contention period of the next frame. The detailed two-phase algorithm of HFCA is depicted in Fig. 6.

Immediately prior to the first phase, the mean number of MTs wishing to contend for contention bandwidth, referred to as the initial group size (denoted as N) is revealed to the BS. This can be realized via *a priori* call distribution or simple prediction. Notice that for the latter case, as will be shown later that HFCA is robust against prediction discrepancy. Without loss of generality, we adopt the use of prediction throughout the rest of the paper. The predicted initial group size $\hat{N}$ is then broadcast to all MTs via the downlink channel.

In the admission phase, each active MT independently accesses the subsequent slot based on the controlled slotted-Aloha (CS-Aloha) protocol parameterized by the admission probability $P_A(\hat{N}) = \kappa_{\text{opt}}/\hat{N}$, where κ_{opt} is the optimal value yielding maximized saturated throughput. As will be formally derived in the throughput analysis, $\kappa_{\text{opt}} \approx 1.52$, irrelevant to the network load. The group being admitted is comprised of all MTs who have actually made transmissions. Upon receiving the composite signal, PMER performs the estimation of the number of colliding MTs ($\hat{L}$). If a collision occurs ($2 \le \hat{L} \le 5$), the operation proceeds to the second phase. All active MTs are, in turn, notified with the $\hat{L}$ value and assigned the value to the reduced group size (m). If $\hat{L} \le 1$, the current two-phase cycle terminates. Another cycle may be initiated subject to the maximum frame boundary limitation. It is worth mentioning that if the number of colliding MTs exceeds five ($\hat{L} > 5$), PMER would fail to perform the estimation. Since no user can be admitted under this case, it is treated the same as the idle case and a feedback of $\hat{L} = 0$ is returned, resulting in only one-slot bandwidth waste. As will be shown later, HFCA incurs only a negligible probability of exceeding five concurrent transmitters.

In the resolution phase, each admitted unresolved MT designates its transmitting probability within the subsequent slot inversely proportional to the current number of MTs in the group. The group size starts with the reduced group size (m) and decremented at most by one per slot. Namely, each MT accesses the next contention slot based on the CS-Aloha protocol parameterized by the resolution probability $P_R(G) = 1/G$, where G is the current group size. As was just mentioned, G is identical to the reduced group size (m) at the beginning of the resolution phase. Notice that the G value is independently maintained at each MT throughout the resolution phase. At the end of slot access, depending on the returned feedback ($\hat{L}$), each MT takes different actions. If $\hat{L} = 1$ (success), the current group size (G) is decremented by one. If $\hat{L} = 0$ (idle), or $\hat{L} \ge 2$ (collision) but $\hat{L} \ne G - 1$, resolution recursively repeats with G remained unchanged. Significantly, if $\hat{L} \ge 2$ (collision) but $\hat{L} = G - 1$, by taking advantage of having a single MT in the nontransmission set, HFCA allows dedicated access within the subsequent slot to this MT. All other MTs in the transmission set then result in a one-slot delay and reduce G by one. The phase-two operation repeats until all users in the admitted group are fully resolved, i.e., $G = 0$.

The HFCA operation is further illustrated via an example given in Fig. 7. In the example, the network starts with 50 stations $\hat{N} = 50$ wishing to transmit signaling requests. In the first phase, each MT independently makes a slot access attempt based on CS-Aloha with probability $P_A = 1.52/50$. Suppose there are four MTs that have actually made transmissions. The number of colliding MTs ($\hat{L} = 4$) is estimated by PMER and broadcast to MTs, triggering the entering of the resolution phase with the reduced group size $m = 4$. In the second phase, each MT performs CS-Aloha with probability $P_R = 1/4$. Suppose

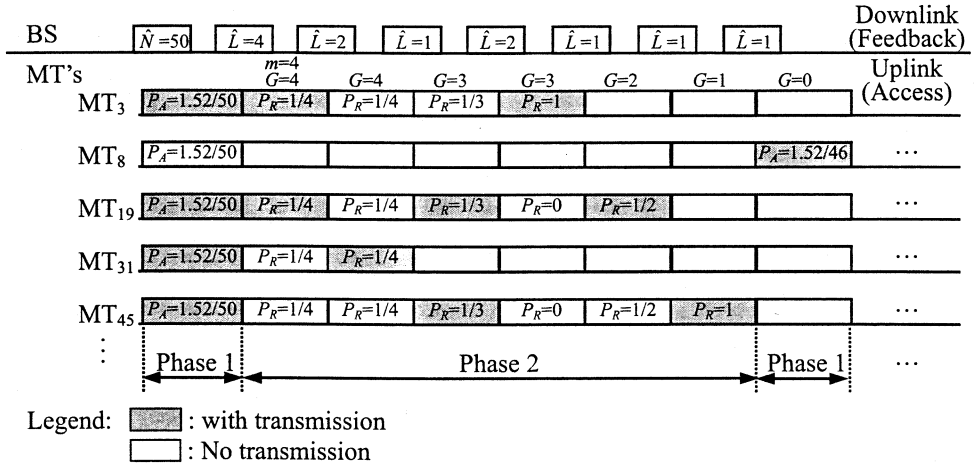


Fig. 7. HFCA operation—an example.

two of them actually make transmissions. The feedback of $\hat{L} = 2$ yields all four MTs to retry on the subsequent slot using the same probability ($P_R = 1/4$). Assuming there is only one MT (MT_{31}) that actually accesses the slot this time. Upon being notified with success access ($\hat{L} = 1$), each MT reduces the group size by one ($G = 3$), and performs CS-Aloha parameterized by $P_R = 1/G = 1/3$. Suppose there are two MTs that actually transmit their requests. The feedback $\hat{L} = 2$ implicitly inhibits the two previous transmitters from accessing the next slot, resulting in dedicated slot access for MT_3. The operation repeats until all four MTs are resolved. At this moment, assuming more requests can be accepted, another two-phase cycle starts over with the initial group size reduced by four, i.e., $\hat{N} = 46$.

B. Throughput and Stability Analyses

Without loss of generality, we discount reservation access and only consider contention access in the following analyses throughout the rest of the section. Each frame contains a contention period comprising repeated two-phase intervals. The system regenerative points are placed at the beginning of each contention period. Let $\tilde{N}$ and N, respectively, represent the random variable and variable for the total number of active MTs wishing to transmit requests at the beginning of the contention period of a frame. In addition, $\tilde{N}$ is assumed Poisson distributed with parameter α, namely $E[\tilde{N}] = \alpha$. We assume that all N users are resolved at the end of the contention period. Denote by random variable $\tilde{C}_N$ the contention period length, namely the total number of slots required to resolve the user group of size N, and let $C_N = E[\tilde{C}_N]$. Let variable m denote the first reduced group size, i.e., the number of active MTs admitted at the end of first Phase 1. Then, denote by random variable $\tilde{D}_m$ the total number of slots required to resolve all m users in a reduced group in Phase 2, and let $D_m = E[\tilde{D}_m]$. Clearly, the value C_N, which is a function of D_m, plays a significant role in the throughput and stability analyses. In the sequel, we first compute C_N (and D_m), followed by throughput and stability analyses.

1) Computation of C_N: If there is no request ($N = 0$), or only a single request ($N = 1$), wishing to transmit in the first slot (first phase) of the contention period, then the contention period lasts for one slot, that is

$$\tilde{C}_0 = \tilde{C}_1 = 1. \tag{10}$$

For $N \geq 2$, the probability that exactly m of N users make transmissions and are admitted in the first phase is

$$\begin{aligned} Q_N(m) &= \binom{N}{m} P_A(N)^m [1 - P_A(N)]^{N-m} \\ &= \binom{N}{m} \left(\frac{\kappa}{N}\right)^m \left(1 - \frac{\kappa}{N}\right)^{N-m}, \quad 0 \leq m \leq N. \end{aligned} \tag{11}$$

Given that m of N users have made transmissions, the conditional length of the contention period can be expressed recursively as

$$\begin{cases} \tilde{C}_{N|m} = 1 + \tilde{D}_m + \tilde{C}_{N-m}, & N > m \geq 2, N \geq 2 \\ C_{N|m} = 1 + \tilde{D}_m, & N = m \geq 2, N \geq 2 \\ \tilde{C}_{N|m} = 1 + \tilde{C}_{N-m}, & m < 2, N \geq 2. \end{cases} \tag{12}$$

The first Term 1 corresponds to the slot for the admission phase. It then takes $\tilde{D}_m$ slots to resolve the reduced group of size m if $m \geq 2$, and $\tilde{C}_{N-m}$ slots to resolve the remaining $N - m$ users if $N - m > 0$. We now compute C_N. First, from (10), we get

$$C_0 = C_1 = 1. \tag{13}$$

For $N \geq 2$, mean C_N can be directly derived by taking expectation from (12), or generally obtained via the z-transform (a moment generation function). Taking expectation on both sides of (12), one gets

$$\begin{cases} C_{N|m} = 1 + D_m + C_{N-m}, & N > m \geq 2, N \geq 2 \\ C_{N|m} = 1 + D_m, & N = m \geq 2, N \geq 2 \\ C_{N|m} = 1 + C_{N-m}, & m < 2, N \geq 2. \end{cases} \tag{14}$$

By unconditioning, it becomes

$$\begin{aligned} C_N &= 1 + \sum_{m=2}^{N} Q_N(m) \cdot D_m + \sum_{m=0}^{N-1} Q_N(m) \cdot C_{N-m} \\ &= 1 + \sum_{m=2}^{N} Q_N(m) \cdot D_m \\ &\quad + \sum_{m=1}^{N-1} Q_N(m) \cdot C_{N-m} + Q_N(0) \cdot C_N, \qquad N \geq 2. \end{aligned} \tag{15}$$

Solving for C_N, we obtain

$$C_N = \frac{1 + \sum_{m=2}^{N} Q_N(m) \cdot D_m + \sum_{m=1}^{N-1} Q_N(N-m) \cdot C_m}{1 - Q_N(0)}, \quad N \geq 2. \tag{16}$$

To show that C_N can also be solved from the z-transform, define $F^*_{\tilde{C}_N}(z) \triangleq E[z^{\tilde{C}_N}]$ and $F^*_{\tilde{D}_m}(z) \triangleq E[z^{\tilde{D}_m}]$. Taking z-transform from both sides of (12) and unconditioning the reduce group size m, we get

$$\begin{aligned} F^*_{\tilde{C}_N}(z) &= \sum_{m=0}^{N} Q_N(m) \cdot E[z^{\tilde{C}_{N|m}}] \\ &= z \cdot \sum_{m=2}^{N-1} Q_N(m) \cdot F^*_{\tilde{D}_m}(z) \cdot F^*_{\tilde{C}_{N-m}}(z) \\ &\quad + z \cdot Q_N(0) \cdot F^*_{\tilde{C}_N}(z) + z \cdot Q_N(1) \cdot F^*_{\tilde{C}_{N-1}}(z) \\ &\quad + z \cdot Q_N(N) \cdot F^*_{\tilde{D}_n}(z). \end{aligned} \tag{17}$$

Solving for $F^*_{\tilde{C}_N}(z)$, we have (18), as shown at the bottom of the page. Taking the first derivative of $F^*_{\tilde{C}_N}(z)$ at $z = 1$, one can easily obtain the same result of (16).

2) Computation of D_m: Given that i of m users ($m \geq 2$) in the reduce group made transmissions in the first slot of the phase-two period, the conditional length of the phase-two period can be formulated recursively as

$$\tilde{D}_{m|i} = \begin{cases} 1 + \tilde{D}_{m-1}, & i = 1 \\ 2 + \tilde{D}_{m-1}, & i = m-1 \\ 1 + \tilde{D}_m, & i = 0, 2 \leq i < m-1, \quad \text{or } i = m. \end{cases} \tag{19}$$

The first and second cases correspond to a successful transmission and dedicated access, respectively. It then takes $\tilde{D}_{m-1}$ slots to resolve the remaining users. The last case then corresponds to idle, or a collision but with more than one user in the nontransmission set. The size of users to be resolved remains unchanged. For $m \geq 2$, the probability that exactly i of m users have made transmissions in the first slot of the Phase 2 period is

$$\begin{aligned} R_m(i) &= \binom{m}{i} P_R(m)^i [1 - P_R(m)]^{m-i} \\ &= \binom{m}{i} \left(\frac{1}{m}\right)^i \left(1 - \frac{1}{m}\right)^{m-i}, \quad m \geq 2. \end{aligned} \tag{20}$$

To compute D_m for $2 \leq m \leq N$, we first have the boundary condition $D_1 = 1$ due to the fact that it requires exactly one slot at the end of Phase 2 if the remaining user size is one. Moreover, the computation of D_m is different under $m > 2$ from that under $m = 2$, owing to the possibility of having dedicated access in the $m > 2$ case. We can solve the second boundary condition for D_2 by the recursive equation: $D_2 = 1 + R_2(1) \cdot D_1 + [1 - R_2(1)] \cdot D_2$. We get $D_2 = 3$. For $m > 2$, taking z-transform from both sides of (19) and unconditioning variable i, we get

$$\begin{aligned} F^*_{\tilde{D}_m}(z) &= \sum_{i=0}^{m} R_m(i) \cdot E[z^{\tilde{D}_{m|i}}] \\ &= E[z^{1+\tilde{D}_m}][1 - R_m(1) - R_m(m-1)] \\ &\quad + E[z^{1+\tilde{D}_{m-1}}] \cdot R_m(1) + E[z^{2+\tilde{D}_{m-1}}] \\ &\quad \cdot R_m(m-1) \\ &= z \cdot F^*_{\tilde{D}_m}(z)[1 - R_m(1) - R_m(m-1)] \\ &\quad + z \cdot F^*_{\tilde{D}_{m-1}}(z) \cdot R_m(1) \\ &\quad + z^2 \cdot F^*_{\tilde{D}_{m-1}}(z) \cdot R_m(m-1), \quad m \geq 3. \end{aligned} \tag{21}$$

Solving for $F^*_{\tilde{D}_m}(z)$, we obtain the recursive form as

$$F^*_{\tilde{D}_m}(z) = \frac{F^*_{\tilde{D}_{m-1}}(z) \cdot [z \cdot R_m(1) + z^2 \cdot R_m(m-1)]}{1 - z \cdot [1 - R_m(1) - R_m(m-1)]}, \quad m \geq 3. \tag{22}$$

Taking the first derivative of $F^*_{\tilde{D}_m}(z)$ at $z = 1$, we obtain

$$\begin{aligned} D_m &= F^{*\prime}_{\tilde{D}_m}(z)\Big|_{z=1} = \frac{1 + R_m(m-1)}{R_m(1) + R_m(m-1)} + D_{m-1} \\ &= \sum_{k=3}^{m} \frac{1 + R_k(k-1)}{R_k(1) + R_k(k-1)} + D_2 + D_1, \quad m \geq 3 \end{aligned} \tag{23}$$

where $D_1 = 1$ and $D_2 = 3$, which have been previously derived.

$$F^*_{\tilde{C}_N}(z) = \frac{z\left(\sum_{m=2}^{N-1} Q_N(m) \cdot F^*_{\tilde{D}_m}(z) \cdot F^*_{\tilde{C}_{N-m}}(z) + Q_N(1) \cdot F^*_{\tilde{C}_{N-1}}(z) + Q_N(N) \cdot F^*_{\tilde{D}_n}(z)\right)}{1 - z \cdot Q_N(0)}, \quad N \geq 2 \tag{18}$$

TABLE II
SYSTEM PARAMETER AND THROUGHPUT UNDER VARIOUS αs

α	10	20	30	40	50	100	150	200	250	300	350	400
κ_{opt}	1.30	1.38	1.42	1.44	1.46	1.49	1.51	1.51	1.52	1.52	1.52	1.52
$\bar{S}_{sat}$	0.554	0.539	0.534	0.531	0.529	0.525	0.524	0.523	0.522	0.522	0.522	0.522
$\bar{S}_{max}$	0.622	0.616	0.613	0.612	0.610	0.607	0.606	0.606	0.605	0.605	0.605	0.605
α_{max}	1.85	1.86	1.87	1.87	1.88	1.88	1.88	1.88	1.88	1.88	1.88	1.88

3) Throughput Computation: Since $\tilde{N}$ is Poisson distributed, with C_N (and D_m) given by (13), (16), and (23), the system throughput can be derived as

$$\bar{S} = \frac{E[\tilde{N}]}{E[C_{\tilde{N}}]} = \frac{\sum_{N=0}^{\infty} N \cdot \frac{e^{-\alpha}\alpha^N}{N!}}{\sum_{N=0}^{\infty} C_N \cdot \frac{e^{-\alpha}\alpha^N}{N!}}. \quad (24)$$

Notice that in (24), system throughput $\bar{S}$ is a function of α and κ. We now formally define the two aforementioned throughputs—*saturated throughput* ($\bar{S}_{\rm sat}$), and *maximum stable throughput* ($\bar{S}_{\rm max}$). Saturated throughput is the throughput achieved when the system is fully saturated (i.e., $\alpha \to \infty$). Namely

$$\bar{S}_{\rm sat} \triangleq \lim_{\alpha\to\infty} \bar{S}. \quad (25)$$

The optimal value $\kappa_{\rm opt}$ applied during the admission phase is chosen such that $\bar{S}_{\rm sat}$ is maximized. That is, $\bar{S}_{\rm sat}(\kappa_{\rm opt}) = \max\{\bar{S}_{\rm sat}(\kappa), \forall\kappa\}$. A system is WSS if it has a positive saturated throughput. Then, maximum stable throughput ($\bar{S}_{\rm max}$) is the maximum achievable throughput when the system is stable.

To numerically evaluate $\kappa_{\rm opt}$, $\bar{S}_{\rm sat}$, and $\bar{S}_{\rm max}$, we carried out analytic experiments via Mathematica 4.0. Prior to the computation, we needed to determine a maximum $N(N_{\rm max})$ to be applied to (24). Since the Poisson distribution converges to the Gaussian distribution with the same mean and variance α, we considered $N_{\rm max} = \alpha + 5\sqrt{\alpha}$, which gives sufficiently high confidence to the evaluation ($P(N > \alpha + 5\sqrt{\alpha}) < 2.866\,53 \times 10^{-7}$). First of all, for a given α, $\kappa_{\rm opt}$ was determined by using Mathematica function $\mathrm{FindMinimum}[1 - \bar{S}[N_{\rm max}, \alpha, \kappa_{\rm opt}], \kappa_{\rm opt}]$. We obtained that $\bar{S}_{\rm sat}$ is maximized under $\kappa_{\rm opt} = 1.52$. Applying $\kappa_{\rm opt}$ to $\bar{S}$ in (25) with $\alpha \geq 400$, we got $\bar{S}_{\rm sat} \approx 0.522$. We conclude that HFCA is WSS. Then, applying the $\kappa_{\rm opt}$ value and using the same function but a different parameter, i.e., $\mathrm{FindMinimum}[1 - \bar{S}[N_{\rm max}, \alpha, 1.52], \alpha]$, we revealed that $\bar{S}$ is maximized at $\alpha_{\rm max} = 1.88$. Applying the $\alpha_{\rm max}$ value to $\bar{S}$ in (24), we arrived at that $\bar{S}_{\rm max} = 0.605$ at $\alpha_{\rm max} = 1.88$. Analytical results under a multitude of αs are summarized in Table II. From the table, we discovered that the $\kappa_{\rm opt}$ values under various αs vary insignificantly under $\alpha \geq 50$, which is usually the case for practical local wireless networks. This justifies the simplicity and feasibility of HFCA in any local environment.

4) Stability Analysis: First of all, as was previously defined, the system is WSS if it has a positive saturated throughput. Thus, based on the previous numerical results ($\bar{S}_{\rm sat} \approx 0.522$), HFCA is WSS. Second, notice that the ratio N/C_N can be perceived as the "effective service rate" [16] of the HFCA system. Then, if there exists a lower bound of N/C_N, then there exists a *system capacity* [17] defined as the supremum of new arrival rates that ensure strict-sense stability of the system.

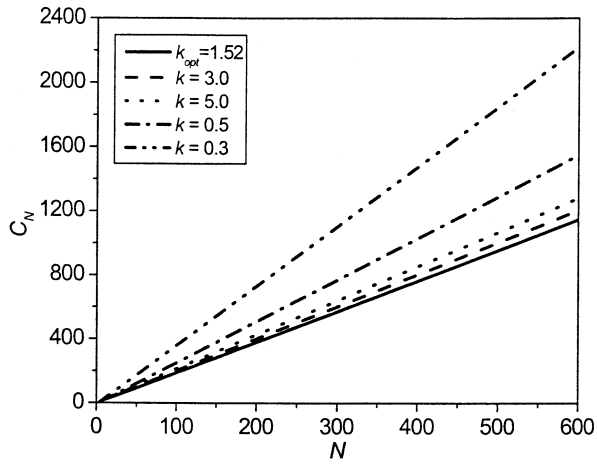


Fig. 8. C_N performance under various κ values.

TABLE III
EFFECTIVE SERVICE RATE FUNCTION (N/C_N)

N	1	2	3	4	5
N/C_N	1	0.608	0.605	0.585	0.574
N	6	7	8	9	10
N/C_N	0.567	0.561	0.557	0.554	0.551
N	20	40	60	80	100
N/C_N	0.538	0.531	0.528	0.526	0.525
N	200	300	400	500	600
N/C_N	0.523	0.522	0.522	0.522	0.522

Due to mathematical intractability for deriving the closed form of C_N from its recursive form in (16), we derived C_N and N/C_N by means of numerical computation. Numerical results of C_N (under various κs and $\kappa_{\rm opt}$) and N/C_N (applying $\kappa_{\rm opt}$) for $N = 1$ to 600 are given in Fig. 8 and Table III, respectively. Based on the numerical evaluations, we claim in the following remark that the system capacity of HFCA is $\bar{S}_{\rm sat}$, i.e., HFCA is SSS if the new arrival rate is lower than $\bar{S}_{\rm sat}$.

Remark: The HFCA system is SSS if the new arrival rate is lower than $\bar{S}_{\rm sat}$.

First, numerical results from Table III show that N/C_N ($N \geq 1$) is a monotonically decreasing function. Next, by taking advantage of linearty of (C_N) shown in Fig. 8, we obtain

$$\bar{S}_{\rm sat} = \lim_{\alpha\to\infty} \frac{E[\tilde{N}]}{E[C_{\tilde{N}}]} = \lim_{\alpha\to\infty} \frac{\alpha}{C_{E[\tilde{N}]}} = \lim_{\alpha\to\infty} \frac{\alpha}{C_\alpha}. \quad (26)$$

Namely, N/C_N converges to $\bar{S}_{\rm sat}$. We can, thus, conclude that N/C_N is lower bounded by $\bar{S}_{\rm sat}$, i.e., $N/C_N \geq \bar{S}_{\rm sat} \approx 0.522$. Therefore, if the new arrival rate is lower than $\bar{S}_{\rm sat}$, it is lower than the effective service rate N/C_N; the system is SSS.

C. Experimental Results

We draw comparisons of performance with respect to throughput, access delay, and blocking probability, between HFCA and four existing schemes via event-based simulation. These schemes are binary-feedback collision resolution (BFCR), Paris [9], clipped binary (CB)-tree [16], and S-Aloha.

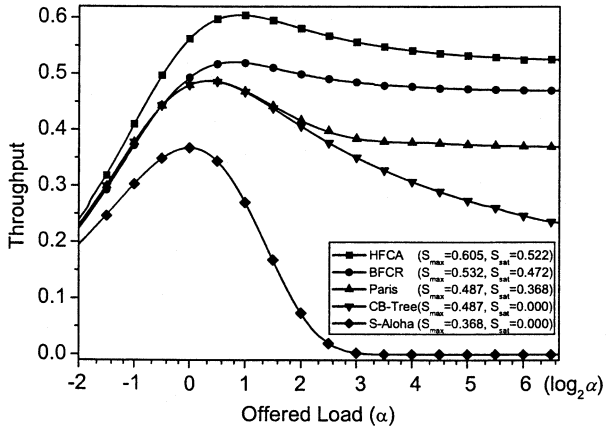


Fig. 9. Comparison of throughput performance.

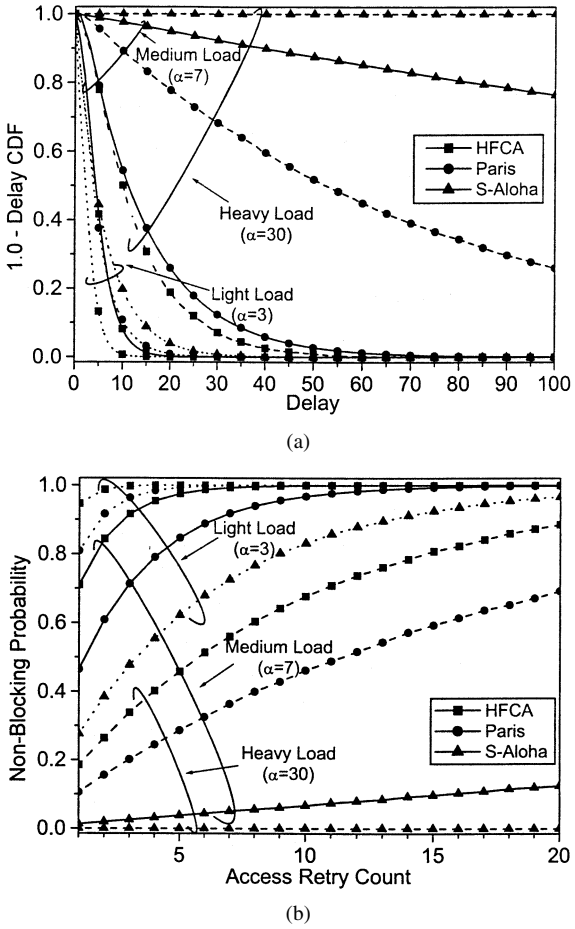


Fig. 10. Comparison of delay and blocking probability. (a) Delay comparison. (b) Nonblocking probability comparison.

The BFCR scheme is the binary-feedback version of HFCA, namely, with PMER removed. Specifically, BFCR exploits the same two-phase collision resolution algorithm and uses binary feedback during the two-phase process. In the simulation, we adopted $P_R = 1/2$ in the resolution phase. For the simulation of the Paris scheme (a near-optimal collision resolution algorithm), we extracted different p values under different offered loads from the $p(u)$ function analytically constructed in advance. For CB-tree, the optimal p value, namely $p = 1/2$, was applied. Simulation was terminated after reaching 95% confidence interval. Simulation results are depicted in Figs. 9 and 10.

In Fig. 9, we discover that HFCA, BFCR, and Paris schemes assure converged saturated throughput, and thus, are WSS. As opposed to them, CB-tree and S-Aloha are inherently unstable rendering a saturated throughput near to zero under heavy loads. In particular, compared to BFCR, HFCA achieves 10% throughput improvement at the cost of simple implementation of PMER. Both BFCR and HFCA invariantly outperform the Paris scheme.

We further make comparisons of access delay and blocking probability among HFCA, Paris, and S-Aloha schemes under low, medium, and high loads ($\alpha = 3,\ 7,$ and 30). Results are plotted in Fig. 10. In the simulation, the access delay of an MT was measured as the total number of slots required until the MT successfully transmits its signaling request. In Fig. 10(a), the complement delay cumulative distribution function is displayed. For monitoring the blocking probability, an MT's request was considered blocked if the number of slot accesses exceeds the predefined access retry count in units of slots. In Fig. 10(b), we show nonblocking probability as a function of the retry count, ranging from 1 to 20. We clearly observe that, owing to being unstable, S-Aloha results in drastic increases in delay and blocking probability under both the medium and heavy loads. Furthermore, HFCA significantly outperforms Paris on both performance metrics particular under medium and heavy loads.

V. System Robustness Assessment

To assess the robustness of the HFCA system, we examine the impact of both prediction and estimation discrepancy on throughput performance. While the former discrepancy is caused by the lack of signaling arrival distribution, the latter is due to estimation error (by PMER). The assessment is based on analytic and simulation results.

First, recall that only if there is no signaling arrival distribution provided should the system require a general prediction method for estimating the initial group size $\hat{N}$. The consequence of incorrect prediction is two-fold. With under-prediction, the system results in an increase in probability of having more than five users making transmission attempts in Phase 1. With over-(and under-) prediction, the system throughput might deteriorate. Denote by random variable $\tilde{n}$, the reduced group size at the end of first phase, and δ, the prediction deviation ratio ($-1 \le \delta \le 1$). With prediction deviation taken into account, admission probability $P_A(\hat{N})$ is replaced by $P_A^E(\hat{N})$. The probability that there are more than five users in a reduced group becomes

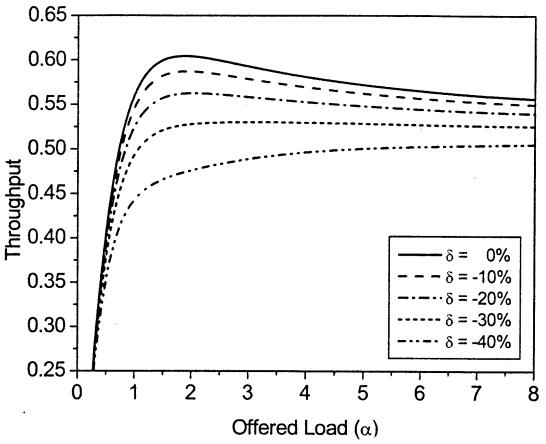
$$p(\tilde{n} > 5) = 1 - \sum_{k=0}^{5} \binom{\hat{N}}{k} P_A^E(\hat{N})^k [1 - P_A^E(\hat{N})]^{\hat{N}-k}$$

$$\text{where } P_A^E(\hat{N}) = \frac{\kappa_{\text{opt}}}{(1+\delta)\cdot\hat{N}}. \quad (27)$$

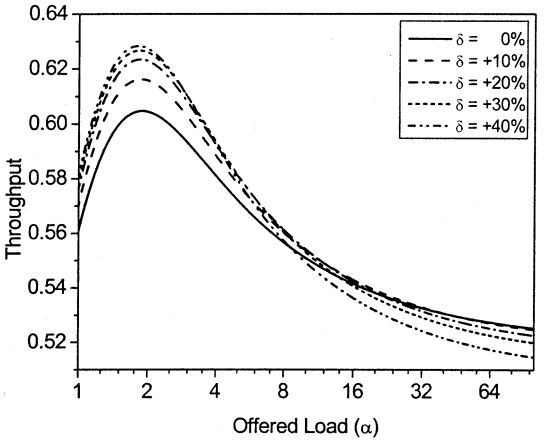
Analytical results under various $\hat{N}$ and δ values are summarized in Table IV. We discover from the table that the probability increase is as insignificant as 0.0511 even under high

TABLE IV
PROBABILITY OF A REDUCED GROUP SIZE GREATER THAN FIVE $p(\bar{n} > 5)$

$\hat{N} = 100$	$\delta = 0\%$	+10%	+20%	+30%	+40%
	0.0052	0.0033	0.0022	0.0015	0.001
		-10%	-20%	-30%	-40%
		0.0086	0.0146	0.0261	0.0491
$\hat{N} = 300$	$\delta = 0\%$	+10%	+20%	+30%	+40%
	0.0055	0.0035	0.0023	0.0015	0.0011
		-10%	-20%	-30%	-40%
		0.009	0.0153	0.0272	0.0507
$\hat{N} = 500$	$\delta = 0\%$	+10%	+20%	+30%	+40%
	0.0056	0.0036	0.0023	0.0016	0.0011
		-10%	-20%	-30%	-40%
		0.0091	0.0155	0.0274	0.0511



(a)



(b)

Fig. 11. Throughput degredation due to prediction discrepancy. (a) Under-prediction of the initial group size. (b) Over-prediction of the initial group size.

δ (-0.4). This fact ensures successful multiuser estimation by PMER throughout the two-phase process.

We further examine the impact of under- and over-prediction on system throughput via simulation. All settings were identical to those in previous experiments, except that $P_A(\hat{N})$ was also replaced by $P_A^E(\hat{N})$, as given in (27). The results are shown in Fig. 11. First, we surprisingly observe in Fig. 11(b) that if the initial group size is over predicted, throughput even improves under light loads and declines insignificantly under heavier loads. This is because κ_{opt} in $P_A(\hat{N})$ has been chosen

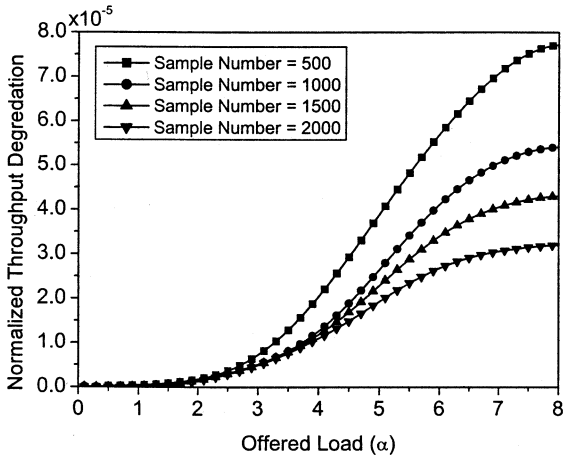


Fig. 12. Normalized throughput degredation due to estimation discrepancy.

to maximize the saturated throughput under heavy loads. Over-prediction produces the same effect as the reduction of the κ value yielding higher maximum stable throughput. For the case of under-prediction, as shown in Fig. 11(a), HFCA only incurs in significant throughput degradation.

We further study the impact of estimation discrepancy on throughput performance. Recall that the use of greater sample numbers by PMER yields higher estimation precision. Based on the estimation precision distributions previously measured and plotted in Fig. 5, we carried out an experiment using four different sample numbers (500, 1000, 1500, and 2000) via simulation. All settings were identical to those from previous simulations, except that the estimation precision probabilities were exerted during multiuser estimation of the two-phase HFCA operation. Simulation results are plotted in Fig. 12. We discover from the figure that under the worst case using 500 samples, throughput deteriorates only by 8×10^{-5}. This justifies the robustness of the system against estimation discrepancy.

VI. CONCLUSION

In this paper, we have proposed a WSS high-performance contention access system, HFCA, and its multiuser estimator, PMER, for supporting signaling traffic over wireless access networks. HFCA is capable of leveraging access efficiency by using a two-phase (admission and resolution) algorithm augmented with hexanary feedback control facilitated by PMER implemented at the physical layer. The two-phase algorithm performs incremental contention resolution, resolving a small subset of users at a time. By means of envelope–phase pdf-based estimation, PMER provides six types of feedback, respectively, associated with six access outcomes during the real-time operation of the two-phase algorithm. With the first-phase parameter κ_{opt} determined in the throughput analysis, HFCA was shown to achieve the highest maximum stable throughput (≈ 0.605) and saturated throughput $\bar{S}_{\text{sat}}$ (≈ 0.522) reported to date. In the stability analysis, we concluded that the system capacity of HFCA is $\bar{S}_{\text{sat}}$, i.e., HFCA is SSS is the new arrival rate is lower than $\bar{S}_{\text{sat}}$. Simulation results demonstrate performance superiority of HFCA in terms of throughput, access delay, and blocking probability. The system was finally

shown to be highly robust against prediction and estimation discrepancy.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments that have greatly improved the quality of the paper.

REFERENCES

[1] Y. Kwok and V. Lau, "A quantitative comparison of multiple access control protocols for wireless ATM," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 796–815, May 2001.

[2] S. Jangi and L. Merakos, "Performance analysis of reservation random access protocols for wireless access network," *IEEE Trans. Commun.*, vol. 42, pp. 1223–1234, Feb./Mar./Apr. 1994.

[3] M. Yuang and P. Tien, "Multiple access control with intelligent bandwidth allocation for wireless ATM networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1658–1669, Sept. 2000.

[4] L. Lenzini, M. Luise, and R. Reggiannini, "CRDA: A collision resolution and dynamic allocation MAC protocol to integrate date and voice in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1153–1163, June 2001.

[5] I. Akyildiz *et al.*, "Medium access control protocols for multimedia traffic in wireless networks," *IEEE Network*, vol. 13, pp. 39–47, July/Aug. 1999.

[6] N. Passas *et al.*, "Quality-of-service-oriented medium access control for wireless ATM networks," *IEEE Commun. Mag.*, vol. 35, pp. 42–50, Nov. 1997.

[7] M. Ivanovich, M. Zukerman, and F. Cameron, "A study of deadlock models for a multiservice medium access protocol employing a slotted aloha signalling channel," *IEEE/ACM Trans. Networking*, vol. 8, pp. 800–811, Dec. 2000.

[8] W. Jeon, D. Jeong, and C. Choi, "An integrated services MAC protocol for local wireless communications," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 352–364, Feb. 1998.

[9] B. Paris and B. Aazhang, "Near-optimum control of multiple-access collision channels," *IEEE Trans. Commun.*, vol. 40, pp. 1298–1309, Aug. 1992.

[10] L. Georgiadis and P. Papantoni-Kazakos, "A collision resolution protocol for random access channels with energy detectors," *IEEE Trans. Commun.*, vol. COM-30, pp. 2413–2420, Nov. 1982.

[11] J. Mosley and P. Humblet, "A class of efficient contention resolution algorithms for multiple access channels," *IEEE Trans. Commun.*, vol. COM-33, pp. 145–151, Feb. 1985.

[12] M. Yuang and B. Lo, "QMAC: A QoS-guaranteed MAC protocol with dynamic granularity control for local wireless ATM networks," *Proc. IEEE Int. Conf. Communications (ICC)*, vol. 8, pp. 2359–2364, June 2001.

[13] Y. Su and J. Chen, "Carrier-to-interference ratio measurement using moments or histograms," *IEEE Trans. Commun.*, vol. 48, pp. 1338–1346, Aug. 2000.

[14] S. O. Rice, "Probability distributions for noise plus several sine waves—The problem of computation," *IEEE Trans. Commun.*, vol. COM-22, pp. 851–853, June 1974.

[15] A. S. Rosenbaum, "PSK error performance with Gaussian noise and interference," *Bell Syst. Tech. J.*, vol. 48, pp. 413–442, Feb. 1969.

[16] R. Rom and M. Sidi, *Multiple Access Protocols-Performance and Analysis*. New York: Springer-Verlag, 1990.

[17] S. Ghez, S. Verdu, and S. Schwartz, "Stability properties of slotted aloha with multipacket reception capability," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 640–649, July 1988.

Maria C. Yuang (M'91–SM'03) received the B.S. degree in applied mathematics from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1978, the M.S. degree in computer science from the University of Maryland, College Park, MD, in 1981, and the Ph.D. degree in electrical engineering and computer science from the Polytechnic University, Brooklyn, NY, in 1989.

From 1981 to 1990, she was with AT&T Bell Laboratories and Bell Communications Research (Bellcore), where she was a Member of the Technical Staff working on high-speed networking and protocol engineering. She was also an Adjunct Professor in the Department of Electrical Engineering, Polytechnic University, Brooklyn, from 1989 to 1990. In 1990, she joined National Chiao Tung University, Hsinchu, Taiwan, R.O.C., where she is currently a Professor in the Department of Computer Science and Information Engineering. Her current research interests include optical and broadband networking, multimedia communications, wireless local/access networking, and performance modeling and analysis.

Bird C. Lo (S'98–M'01) was born in Hsinchu, Taiwan, R.O.C., in 1971. He received the B.S. and M.S. degrees in computer science and engineering from the Yuan Zue University, Chungli, Taiwan, R.O.C., in 1993 and 1996, respectively, and the Ph.D. degree in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2002.

His current research interests include broadband and wireless networking, multimedia communications, performance modeling and analysis.

Ju-Ya Chen (S'96–M'01) received the B.S., M.S., and Ph.D. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1992, 1994, and 2000, respectively.

He was with the Computer and Communications Laboratories of the Industrial Technical Research Institute, where he was involved in the design of mobile communication systems. Since 2001, he has been with the Institute of Communication Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. His research interests include statistical signal processing and spread spectrum communications.