# Back-of-the-Envelope Computation of Throughput Distributions in CSMA Wireless Networks

Soung Chang Liew, *Senior Member*, *IEEE*, Cai Hong Kai, Hang Ching (Jason) Leung, and Piu (Bill) Wong

**Abstract**—This work started out with our discovery of a pattern of throughput distributions among links in IEEE 802.11 networks from experimental results. This pattern gives rise to an easy computation method, which we term back-of-the-envelop (BoE) computation. For many network configurations, very accurate results can be obtained by BoE within minutes, if not seconds, by simple hand computation. This allows us to make shortcuts in performance evaluation, bypassing complicated stochastic analysis. To explain BoE, we construct a theory based on the model of an "ideal CSMA network" (ICN). The BoE computation method emerges from ICN when we take the limit $c \to 0$, where $c$ is the ratio of the mean backoff countdown time to the mean transmission time in the CSMA protocol. Importantly, we derive a new mathematical result: the link throughputs of ICN are insensitive to the distributions of the backoff countdown time and transmission time (packet duration) given the ratio of their means $c$. This insensitivity result explains why BoE works so well for practical 802.11 networks, in which the backoff countdown process is one that has memory, and in which the packet size can be arbitrarily distributed. Our results indicate that BoE is a good approximation technique for modest-size networks such as those typically seen in 802.11 deployments. Beyond explaining BoE, the theoretical framework of ICN is also a foundation for fundamental understanding of very-large-scale CSMA networks. In particular, ICN is similar to the Ising model in statistical physics used to explain phenomena arising out of the interactions of a large number of entities. Many new research directions arise out of the ICN model.

**Index Terms**—Access schemes, algorithm/protocol design and analysis, protocols and wireless.

## 1 Introduction

This paper concerns the computation of the throughput distributions of links in carrier-sense multiple-access (CSMA) wireless networks, such as the IEEE 802.11 networks. While methods for throughput computation of CSMA networks now appear in standard textbooks, most, if not all, deal with the case in which each and every link can sense all other links. That is, carrier sensing is all inclusive.

With their widespread deployment, it is now common to find numerous 802.11 networks colocated in the neighborhood of each other. The carrier-sensing relationships among the links of these networks are **non-all-inclusive** in that each link may only sense a subset, but not all, of other links. Compounding the situation is that each link may sense a different subset of links. It is extremely difficult to extend the analytical methods for all-inclusive carrier-sense networks (e.g., [1]) to the non-all-inclusive case because of the inhomogeneity of the state spaces of the links. Good approximation techniques for engineering purposes are highly desirable.

This paper proposes a simple approximation technique which we refer to as a back-of-the-envelop (**BoE**) method. For networks of modest size, the results can be obtained in a matter of minutes, if not seconds, by simple hand computation. A practical application of BoE is for quick identification of performance problems (e.g., unfair throughput distributions and starvations among links) in a network so that remedies could be devised. In particular, BoE allows us to make shortcuts in system performance evaluation.

Section 2 will specify the BoE method formally. To illustrate its simplicity, let us first describe its mechanic without justification. Consider the network in Fig. 1a. Suppose that the links in the network are greedy in that they always have packets to send. An interesting question is as follows: competing among each other under the CSMA medium access control (MAC) protocol, what will be the throughput of each of the link?

In a CSMA network, a link will not initiate a transmission if it senses the transmission of another link within its carrier-sensing range (CSRange). The carrier-sensing relationships among links are described by the contention graph in Fig. 1b in which links are represented by vertices; and an edge joins two vertices if the transmitters of the two associated links can sense each other. The "normalized" throughputs of the links $(Th_1\ Th_2\ Th_3\ Th_4)$ can be quickly approximated to be $(1\ 0\ 0.5\ 0.5)$ within seconds, as described below.

With respect to the contention graph, we try to put a label of 1 to as many vertices as possible with the constraint that two vertices joined by an edge cannot both

- *S.C. Liew and C.H. Kai are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: {soung, chkai6}@ie.cuhk.edu.hk.*
- *H.C. Leung and P. Wong are with Altai Technologies Limited, Unit 209, 2/F, Lakeside 2, 10 Science Park West Avenue, HK Science Park, Shatin, Hong Kong. E-mail: {hcleung, billwong}@altaitechnologies.com.*

*Manuscript received 23 Sept. 2008; revised 4 Sept. 2009; accepted 15 Dec. 2009; published online 6 May 2010.*
*For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2009-02-0040. Digital Object Identifier no. 10.1109/TMC.2010.89.*

1536-1233/10/$26.00 © 2010 IEEE Published by the IEEE CS, CASS, ComSoc, IES, & SPS

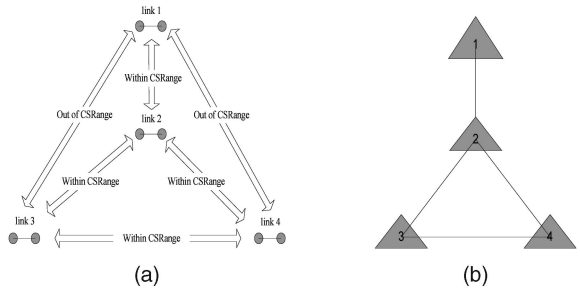



Fig. 1. (a) An example network. (b) Its associated contention graph.

be 1 (i.e., 1 represents transmission, and two links that can sense each other cannot transmit together). We identify two such possibilities: (1 0 1 0) and (1 0 0 1). We then add the vectors together and divide the sum by 2, yielding $(1\ 0\ 0.5\ 0.5)$. And these are the normalized throughputs $(Th_1\ Th_2\ Th_3\ Th_4)$. The procedure is purely algorithmic, and there is no complex stochastic analysis.

For networks of up to 10 links, BoE results can generally be obtained by hand computation; for networks of up to 100 links, we have found that results can be calculated by a computer program within a few minutes. When there are more than 100 links, the computation time of BoE can begin to get out of hand. For infrastructure networks such as those deployed in a typical building (to date, most 802.11 networks are deployed in the infrastructure mode), we propose a **link aggregation** technique that can allow us to handle even more links. We, therefore, believe that BoE is a good approximation technique for modest-size networks such as those typically seen in practice.

Historically, we were led to the BoE algorithm by observation of simulation and real-network experimental results rather than by theoretical construction.[1] A theory was developed thereafter to explain the experimental observation. In this paper, we attempt to explain BoE with **an ideal CSMA network model (ICN)**. In ICN, at any given time, a link can be in either the backoff state or the transmission state. BoE emerges from ICN when the ratio of the mean backoff countdown time to the mean transmission time (packet duration), $c$, approaches zero. The "small $c$" approximation turns out to be highly accurate for many network topologies, as will be seen. Importantly, we derive a new mathematical result that the link throughputs in ICN are insensitive to the detailed distributions of the backoff countdown time and transmission time given $c$. This **insensitivity result** explains why BoE works so well for 802.11 networks, in which the backoff countdown process is one that has memory, and in which the packet size can be arbitrarily distributed in practice.

Beyond explaining BoE, the theoretical framework of ICN is a foundation for fundamental understanding of very-large-scale CSMA networks. In particular, ICN is similar to the Ising model in statistical physics used to explain phenomena arising out of the interactions of a large number of physical entities (e.g., magnetic spins). Section 5 of this paper will discuss several interesting research directions arising out of this model.

### 1.1 Related Work

There have been numerous publications on non-all-inclusive carrier-sense networks and this is indeed a "hot topic" among researchers. Recent work includes [2], [3], [4], [5], [6], [7], from which earlier work can be traced. Most of the prior methods are stochastic analytical and not algorithmic in nature. BoE, in this paper, is algorithmic and is simpler. The Appendix of our technical report [8] compares BoE with some of these prior methods.

As far as we know, the CSMA network model with exponential idle and transmission times was first considered in [9]. More recent work that models the backoff and transmission processes with exponential distributions includes [10], [11]. This model corresponds to a subclass of ICN. The assumption of exponential backoff time, however, is not compatible with practical CSMA protocols (e.g., 802.11).

In practical CSMA protocols, the backoff process is controlled by a counter and has "memory." Starting with an initial counter value, the counter is decremented incrementally with time; when its value reaches zero, transmission begins. The countdown process, however, freezes whenever it hears a neighbor start to transmit. When the transmission of the neighbor completes, the countdown resumes with the previous counter value when the link was last frozen. In this sense, the countdown process is nonmemoryless.

The general ICN model in this paper takes this memory effect into account. An important contribution of this paper is our proof that the link throughputs in ICN are **insensitive to this memory effect and to the detailed distributions of the backoff countdown and transmission times given the ratio of their means** $c$. The practical significance of this result is that we do not have to operate a CSMA network using a memoryless exponential countdown process or artificially require the packet size to be exponentially distributed (an impractical proposition) in order to have essentially the same link-throughput results.[2] In particular, this result expands the scope of the application of ICN and BoE.

ICN can be cast into the framework of the loss circuit-switched network described in [12], [13]. In the simplest form of the loss network, the interarrival and holding times of calls are exponentially distributed, and this corresponds to the ICN in which the backoff countdown and transmission times are exponential, respectively. It is also known that the steady-state probability distribution of loss networks is insensitive to the holding-time distribution [13]. The correspondence between the two models breaks down, however, when we consider the fact that countdown time in ICN is not exponentially distributed in general and has memory. In the context of the loss circuit-switched network, our insensitivity proof of ICN can be considered as a generalization of the known results of the loss network: in particular, the exponential interarrival time is not essential to the system stationary distribution and can be relaxed.

1. Many real-network experiments were conducted during product development at Altai Technologies. The throughput patterns that give rise to BoE were first observed from these experiments.

2. Note that while the link throughputs are insensitive to the distribution of countdown time, the variance of the delay experienced by a link can be reduced with countdown with memory. This is a reason why, in practical, CSMA networks (e.g., IEEE 802.11) countdown with memory is adopted.

It turns out that the system-state probability distribution of ICN is a **Markov random field** (MRF) [14], [15] and is related to the Ising model in statistical physics. The Appendix of this paper provides a proof for the MRF property.

### 1.2 Paper Organization

The remainder of this paper is organized as follows: Section 2 states the BoE computation method formally and provides simulation and real-network experimental results to demonstrate its accuracy. Section 3 is devoted to the construction of ICN and the explanation of BoE in terms of ICN. Section 4 proves that the link throughputs of ICN are insensitive to the distributions of the backoff and transmission times given the ratio of their mean. Section 5 discusses the implications, applications, and extensions of BoE and ICN. Section 6 concludes this paper.

## 2 BoE Computation Procedure and Experimental Verification

This section focuses on the BoE computation procedure and experimental verification of its accuracy. We also provide a link aggregation technique to allow BoE to handle large-size networks deployed in the infrastructure mode. To keep our focus, we defer the theoretical explanation of why BoE works to Sections 3 and 4. Readers who are more interested in theory could jump straight to Sections 3 and 4 after the formal specification of BoE in the next two paragraphs, and return to the experimental results here later.

The formal description of the BoE computation procedure is given as follows:

**BoE Computation.**

1. Draw the contention graph of the network.
2. Identify the maximum independent sets (MISs) of the contention graph.
3. The normalized throughput of link $i$ is $n_i/n$, where $n$ is the number of MISs identified in Step 2 and $n_i$ is the number of MISs in which link $i$ appears.
4. Convert normalized throughputs to throughputs in bps.

An independent set (IS) is a subset of vertices such that no edge joins any two of them; and an MIS is an IS with the maximum cardinality [16]. We note that counting MIS is an NP-complete problem, and therefore, BoE can get out of hand for networks of very large size. However, for networks of modest size, such as 802.11 networks within a building, the problem is manageable. Generally, for networks of up to 10 links, hand computation is possible and quick. By contrast, even for networks of less than 10 links, inhomogeneous[3] stochastic analysis that extends the framework of the homogeneous stochastic analysis in [1] can be rather complicated. For random networks of up to 100 links, a MATLAB program implementing the BoE computation method can yield results within a few minutes. In addition, the link aggregation technique discussed at the end of this section can be used to alleviate the computation problem in even larger networks.

3. The inhomogeneity is due to different links sensing different subsets of other links. As a result, the state-space reduction decomposition method propounded in [1], which requires all links to have the same experience, is not applicable anymore.

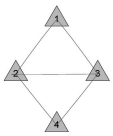
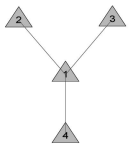
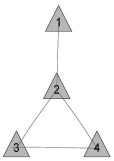
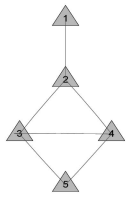
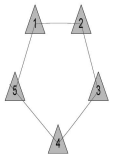
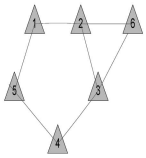
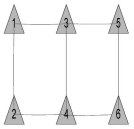
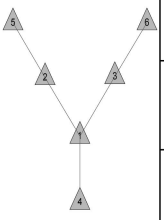
| | | | | |
|---|---|---|---|---|
| (1)  | BoE | (1, 0, 0,1) | (2)  | (0, 1, 1, 1) |
| | NS2, UDP | (0.96, 0.02, 0.02, 0.97) | | (0.03, 0.99, 0.98,1) |
| | NS2, TCP | (0.97, 0.01, 0.01, 0.97) | | (0.02, 0.99, 1, 1) |
| | Real network | (0.76, 0.10, 0.10, 0.74) | | (0.1, 0.8, 0.8, 0.8) |
| (3)  | BoE | (1, 0, 0.5, 0.5) | (4)  | (0.75, 0.25, 0.25, 0.25, 0.5) |
| | NS2, UDP | (0.99, 0, 0.50, 0.51) | | (0.75, 0.26, 0.26, 0.26, 0.51) |
| | NS2, TCP | (0.99, 0, 0.49, 0.51) | | (0.74, 0.25, 0.24, 0.25, 0.50) |
| (5)  | BoE | (0.4, 0.4, 0.4, 0.4, 0.4) | (6)  | (1, 0, 0, 1, 0, 1) |
| | NS2, UDP | (0.41, 0.40, 0.41, 0.40, 0.40) | | (0.92, 0.04, 0.04, 0.92, 0.04, 0.93) |
| | NS2, TCP | (0.40, 0.39, 0.40, 0.40, 0.40) | | (0.95, 0.02, 0.02, 0.95, 0.03, 0.95) |
| (7)  | BoE | (0.5, 0.5, 0.5, 0.5, 0.5, 0.5) | (8)  | (0.2, 0.4, 0.4, 0.8, 0.6, 0.6) |
| | NS2, UDP | (0.5, 0.5, 0.49, 0.5, 0.5, 0.5) | | (0.20, 0.41, 0.41, 0.80, 0.60, 0.60) |
| | NS2, TCP | (0.46, 0.51, 0.51, 0.46, 0.46, 0.51) | | (0.21, 0.39, 0.4, 0.79, 0.61, 0.60) |

Fig. 2. Contention graphs of various network topologies and the corresponding BoE-computed results and experimental results.

Refer to Fig. 1 again. The MISs identified in Step 2 are (1 0 0 1) and (1 0 1 0). Step 3 determines the normalized throughput distribution to be (1 0 0.5 0.5). In Step 4, a normalized throughput of 1 corresponds to the throughput of a link transmitting in isolation of other links as if it were the only link in the whole network. After taking into account the various header, backoff, and ACK overheads (see Section 3.1), the throughput of an isolated link for a UDP session is

$$Th_{\text{single link}} = \begin{array}{l} 6.06 \text{ Mbps (for 11 Mbps 802.11b)} \\ 25.38 \text{ Mbps (for 54 Mbps 802.11g)} \\ 29.45 \text{ Mbps (for 54 Mbps 802.11a).} \end{array}$$

The actual throughput of a link with a normalized throughput of $Th_{\text{norm}}$ is then computed as

$$Th_{\text{actual}} = Th_{\text{norm}} \cdot Th_{\text{single link}}.$$

Fig. 2 shows the results of BoE computation for various network topologies, as well as the corresponding NS2 simulations results for UDP and TCP sessions. Typical

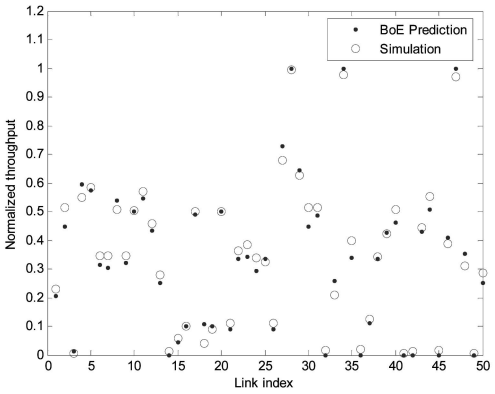



Fig. 3. Verification of BoE for a random network of 50 links.

802.11b and default NS2 parameters were used in the simulations:

1. data rate and basic rate of 11 and 1 Mbps, respectively;
2. carrier-sensing range of 550 m;
3. packet payload of 1,460 Bytes for both UDP and TCP;
4. for each UDP session, a CBR flow of $7\ \text{Mbps} > Th_{\text{single link}} = 6.06\ \text{Mbps}$ was used to ensure link saturation; TCP, being greedy in nature, automatically ensures link saturation; after taking into account TCP_ACK, $Th_{\text{single link}} = 4.84\ \text{Mbps}$.
5. The link length was fixed to 5 m for all links. Each simulation simulated 200 seconds of network dynamic.

As can be seen in Fig. 2, the accuracy of BoE is quite amazing for such a simple method. BoE is also borne out by simulations of numerous other topologies of similar size not shown here.

For larger networks, we ran 10 experiments with 10 randomly generated 50-link networks. The link density is 5 links per square kilometer. The mean degree of links (number of neighbors per link) is around 4. For each link, we calculated the error of the throughput computed by BoE relative to the simulated throughput. The error is normalized by the maximum link throughput in the network. The average link throughput error of the 10 runs is around 5 percent. Fig. 3 shows the detailed comparison of one such run, where the normalized throughput of each link is plotted. As can be seen, BoE is a highly accurate estimation method. More BoE results will be presented in Section 3.5.

Besides simulations, BoE is also borne out by many real-network experiments performed by us. We present the results of two experiments here. We set up two topologies ((1) and (2) in Fig. 2) with four pairs of DELL Latitude D505 laptops PCs with 1.5 GHz Celeron Mobile CPU. Each node has a NETGEAR WAG511GE Dual Band Wireless PC card, and run Fedora5 with MADWifi [17] driver. All Atheros chipset extensions are disabled. The experiments were conducted outdoor on 802.11a channel 36. The throughput of an isolated link is around 29 Mbps. As shown in Fig. 2, the experiment results match well with BoE prediction. In the real environment, we found it difficult to totally isolate two links to keep them out of carrier-sensing range. This is the reason why the measured throughput distributions are not as extreme as predicted.

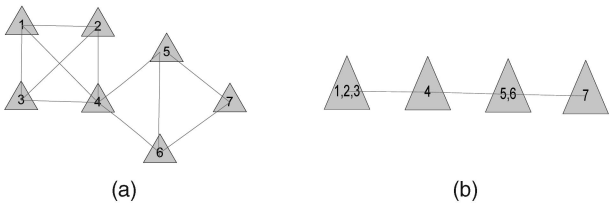



Fig. 4. Illustration of link aggregation. (a) The original contention graph. (b) The simplified link-aggregated contention graph.

**BoE with Link Aggregation.** These days, inside a building, there are typically many 802.11 WLANs deployed in the infrastructure mode in which one end of a link is always an access point (AP). Consider the links associated with a particular AP. It is likely that some of these links will form a clique (i.e., they all sense each other) and they have the same carrier-sensing relationships with other links (i.e., they also sense the same set of other links in the neighborhood). In such a setting, a technique of link aggregation can be used to allow BoE to handle much larger networks. We describe the mechanic of BoE with link aggregation in the next paragraph, with the aid of the example in Fig. 4. We omit detailed justification here that the procedure described below will yield the same results as the original BoE is self-evident after a little thought.

In Fig. 4, the contention graph in Fig. 4a becomes the "virtual" contention graph in Fig. 4b after link aggregation. The virtual contention graph is a linear graph with four virtual vertices {1,2,3}, {4}, {5,6}, {7}. Note that links 1, 2, and 3 sense the same link 4, and so, they can be aggregated. Links 5 and 6 also sense the common links 4 and 7, and so, they can also be aggregated.

The virtual MISs for the virtual contention graph are (1 0 1 0), (1 0 0 1), and (0 1 0 1). Let us define the cardinality of a virtual vertex as the number of original links it contains (e.g., the cardinality of {1,2,3} is 3); and the cardinality of a virtual MIS as the product of the cardinalities of the virtual vertices it contains (e.g., the cardinalities of (1 0 1 0), (1 0 0 1), and (0 1 0 1) are $3 \cdot 2 = 6$, $3 \cdot 1 = 3$, and $1 \cdot 1 = 1$, respectively). Note that each virtual MIS can be mapped to one or more original MIS, and the cardinality of a virtual MIS is the number of original MIS it represents. Thus, the total number of MISs, $n$, in the original contention graph is the sum of the cardinalities of the virtual MIS, which, in our example, is $6 + 3 + 1 = 10$. The normalized throughput distribution of the virtual links is $\sum_k (\text{MIS}_k \times \text{Cardinality of MIS}_k)/n$. In our example, we have $[(1\ 0\ 1\ 0) \times 6 + (1\ 0\ 0\ 1) \times 3 + (0\ 1\ 0\ 1)]/10 = (0.9\ 0.1\ 0.6\ 0.4)$. The throughput of an original link is then the throughput of its virtual link divided by the cardinality of its virtual link. In our example, links 1, 2, and 3 each has a throughput of $0.9/3 = 0.3$; link 4 has a throughput of 0.1; links 5 and 6 each has a throughput of $0.6/2 = 0.3$; and link 7 has a throughput of 0.4.

As mentioned earlier, we were originally led to BoE from observation of simulation and real-network experimental results rather than from analytical construction. The structural simplicity of BoE led us to believe that there

might be a deeper underlying theory. Sections 3 and 4 detail our attempts to unveil the secret behind BoE.

## 3 Explaining BoE Using ICN

In this section, we attempt to explain BoE in terms of an ICN model. Not long after BoE was discovered from experimental observations, we quickly realized two essential underpinnings of Steps 1-3 of BoE (see Section 2), as embodied in the following propositions:

**Proposition 1.** *The system spends most of its time in MIS, and very little time in other states.*

**Proposition 2.** *The MIS states are equally likely. That is, the system spends approximately equal amounts of time in each MIS.*

Step 2 of BoE makes the rough approximation that the system spends zero time in non-MIS. Step 3 of BoE implicitly assumes Proposition 2. We will show that Proposition 2 is implied by ICN, and Proposition 1 is obtained in the limit of $c \to 0$ in ICN. Our experiments indicate that taking this limit results in very good approximations as far as the typical $c$ used in 802.11 networks is concerned. More generally, we could envision a system in which different links adopt different values of $c$. In this case, some non-MIS may also have appreciable probabilities and cannot be ignored. Extensions along this direction as a resource allocation problem will be discussed in Section 5.

Before we begin, let us briefly review the 802.11 CSMA protocol [18] to relate the various parameters in 802.11 to the BoE and ICN models.

### 3.1 Quick Review of 802.11 CSMA Protocol

In 802.11, a station that has packets to send must first sense the channel to be idle for duration of Distributed Interframe Spacing (DIFS) plus a random number of backoff timeslots before transmitting a packet. For each new transmission attempt, the station chooses a random integer backoff countervalue uniformly distributed in the range of [0, CW], where CW is referred to as the contention window. For a new packet with no prior collisions, CW is initially set to CWmin. The backoff counter is decremented by one for each slot the channel is sensed idle. If the channel is sensed busy before the counter reaches zero, the countdown is frozen until the channel is sensed idle for a DIFS period again, whereupon the countdown continues with the **previous** countervalue when it was last frozen. After transmitting a packet, the sender expects to receive an acknowledgement (ACK) after a Short Interframe Spacing (SIFS) period.

For an isolated link, the time consumed by a successful packet transmission consists of

1. PACKET duration consisting of physical-layer preamble/header, MAC Header, and data payload,
2. SIFS,
3. ACK,
4. DIFS, and
5. the random number of backoff countdown timeslots.

For each packet, the airtime within its carrier-sensing range that must be exclusively dedicated to it is

$$T_{tr} = \text{PACKET} + \text{SIFS} + \text{ACK} + \text{DIFS}. \quad (1)$$

In addition, it also consumes a random backoff countdown time of $T_{cd}$ (i.e., component 5 above). In this paper, we define the countdown overhead as

$$c = E[T_{cd}]/E[T_{tr}]. \quad (2)$$

When collisions are rare, $E[T_{cd}] \approx T_{slot} \times CW_{\min}/2$. Note that $T_{cd}$ is the "active" countdown time of a link and frozen time due to a neighbor's transmission is not included in it. Thus, $c$ does not include the freezing overhead.

For brevity, henceforth, in this paper, we will refer to the backoff countdown process as the countdown process. In addition, the terms "backoff" and "countdown" will be used interchangeably. $T_{tr}$ will be referred to as the transmission time, and $T_{cd}$ will be referred to as the countdown time. We will allow $T_{tr}$ and $T_{cd}$ to have arbitrary probability distributions in our general treatment of ICN.

### 3.2 Definition of ICN

**Definition of ICN.** In ICN, the countdown times of links are modeled as **continuous** random variables. The distributions of countdown time $T_{cd}$ and transmission time $T_{tr}$ can be arbitrary otherwise. When a link completes a transmission, it begins countdown with $T_{cd}$ generated according to the probability density $f(t_{cd})$. When the countdown completes, the link begins to transmit with $T_{tr}$ generated according to the probability density $g(t_{tr})$.[4] The countdown of a link is frozen whenever at least one of its neighbors transmits, and the remaining countdown time $RC$ will be stored; when all neighbors stop transmitting, the countdown will resume with the previously stored $RC$.

Let $S_i \in \{0, 1\}$ denote the state of link $i$, where $S_i = 1$ if link $i$ is transmitting and $S_i = 0$ if link $i$ is not transmitting. When link $i$ is not transmitting, it is either actively counting down or frozen. We can define the **system state** of an ICN with $L$ links as $S = S_1 S_2 \ldots S_L$, where the feasible states correspond to the independent sets of the contention graph of the network. Note that if links $i$ and $j$ are neighbors, then the states in which $S_i$ and $S_j$ are both 1 are not allowed because: 1) they can sense each other and 2) the probability of them counting down to zero and transmitting together is 0 under ICN (because the countdown time is a continuous random variable).

Fig. 5 shows the state-transition diagram of the network in Fig. 1 under the ICN model. To avoid clutters, in Fig. 5, we have merged the two directional transitions between two states into one line. Each transition from left to right corresponds to the beginning of a transmission on one particular link, while the reverse transition corresponds to the ending of a transmission on the same link. For example, the transition from 1000 to 1010 is due to the beginning of a transmission on link 3 while link 1 is transmitting; while the reverse transition from 1010 to 1000 is due to the ending of a transmission on link 3 while the transmission of link 1 continues.

4. We assume that all links have the same countdown and transmission time distributions here. The methods and results in this paper can easily be extended to the general case where different links have different distributions. Also, while we require $T_{cd}$ to be continuous, $T_{tr}$ could be discrete (i.e., $g(t_{tr})$ could consist of a sum of delta functions).

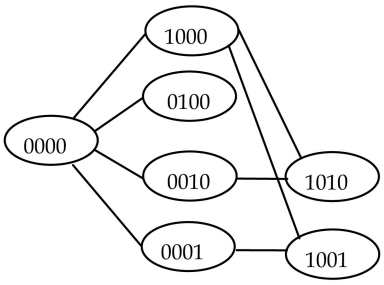



Fig. 5. The state-transition diagram of the network in Fig. 1 under ICN model.

### 3.3 ICN with Memoryless Exponentially Distributed Backoff Time and Transmission Time

According to the definition of ICN, the distributions of $T_{cd}$ and $T_{tr}$ can be arbitrary. To simplify exposition, however, we first assume that their distributions are exponential. In Section 4, we will show that the results obtained here remain valid even if their distributions are not exponential.

**Definition of state connectivity.** Two feasible state realizations $S = s$ and $S = s'$ are said to be connected if it is possible to have a direct transition from $s$ to $s'$, and vice versa, without traversing other states. For example, in Fig. 5, 1000 and 1010 are connected; but 1010 and 1001 are not connected.

**Observation 1.** *Two feasible states $s$ and $s'$ are connected if and only if all the links transmitting in $s$ are also transmitting in $s'$, and there is one extra link transmitting in $s'$ that is not transmitting in $s$, or vice versa.*

**Definition of left and right states.** For two connected states, we refer to the state with one fewer (more) link transmitting as the left (right) state.

**Throughput Computation in ICN.**

Let $P_s = P_{s_1 s_2 \ldots s_L}$ be the fraction of time the network is in a particular state $s = s_1 s_2 \ldots s_L$. Under the exponential distribution assumption, $S(t)$ is a Markov process. The fraction of time link $i$ is transmitting is $x_i = \sum_{s:s_i=1} P_{s_1 s_2 \ldots s_L}$, which corresponds to the normalized throughput of link $i$.

Let $1/\lambda = E[T_{cd}]$ and $1/\mu = E[T_{tr}]$. Then, for any pair of connected states, the transition from the left state to the right state occurs at rate $\lambda = 1/E[T_{cd}]$, and the transition from the right state to the left state occurs at rate $\mu = 1/E[T_{tr}]$. It is easy to verify that the resulting continuous-time Markov chain is time reversible, and therefore, detailed balance applies [15]. Specifically, for two connected states $s$ and $s'$, with $s$ being the left state and $s'$ being the right state, we have

$$P_s = cP_{s'}, \tag{3}$$

where $c = E[T_{cd}]/E[T_{tr}] = \mu/\lambda$ is the countdown overhead defined in (2).

An immediate corollary of Observation 1 and (3) is that all feasible states with the same number of transmitting links (i.e., states in the same column of the state-transition diagram) have the same probability. Specifically, let $S^{(n)}$ be the subset of feasible states with $n$ transmitting links. Then,

$$P_s = \frac{B}{c^n} \forall s \in S^{(n)}, \text{ where } B = \left( \sum_{n=0}^{L} \frac{|S^{(n)}|}{c^n} \right)^{-1}. \tag{4}$$

Applying (4) to the state-transition diagram of Fig. 5 gives

$$\begin{aligned} P_{0000} &= B = \left(1 + 4/c + 2/c^2\right)^{-1} \\ P_{1000} &= P_{0100} = P_{0010} = P_{0001} = (c + 4 + 2/c)^{-1} \\ P_{1010} &= P_{1001} = \left(c^2 + 4c + 2\right)^{-1}. \end{aligned} \tag{5}$$

The normalized throughputs of the links are then given by

$$\begin{aligned} x_1 &= P_{1000} + P_{1010} + P_{1001} = (c + 4 + 2/c)^{-1} + 2\left(c^2 + 4c + 2\right)^{-1} \\ x_2 &= P_{0100} = (c + 4 + 2/c)^{-1}, \\ x_3 &= P_{0010} + P_{1010} = (c + 4 + 2/c)^{-1} + \left(c^2 + 4c + 2\right)^{-1} \\ x_4 &= P_{0001} + P_{1001} = (c + 4 + 2/c)^{-1} + \left(c^2 + 4c + 2\right)^{-1}. \end{aligned} \tag{6}$$

### 3.4 Mapping ICN Results to BoE Computation Method

Steps 1-3 of BoE are implied by the ICN results in (4) in the limit that $c \to 0$. Recall from the beginning of Section 3 that BoE is founded on Propositions 1 and 2. According to (4), in the limit $c \to 0$, only the MIS (the rightmost states in the state-transition diagram) has significant probabilities, and the probabilities of the other IS in ICN become negligible compared with the probabilities of MIS because of the factor $1/c^n$ . In addition, all the MISs have equal probability. In short, Proposition 1 is obtained when $c \to 0$, and Proposition 2 is due to the form of the ICN results in (4).

**Converting Normalized Throughputs to Actual Throughputs.** In the final step of BoE (Step 4), we need to convert the normalized throughputs to actual throughputs in bits per second. This is the engineering part with different alternatives. This paper adopts a very simple procedure. We simply multiply the normalized throughputs by the raw throughput of a single isolated link to obtain the throughputs in bits per second (see Section 2). We note the following: 1) this procedure may underestimate the throughputs because in the single isolated link case, the countdown time is not shared, whereas in the multiple-link case, the countdowns of different links (even if they are neighbors) may occur concurrently. 2) Collisions of links are ignored and this may lead to overestimation of throughputs. Thus, 1) and 2) have opposing effects that pleasingly cancel each other somewhat. Of course, more sophisticated perturbation techniques could be used to adjust for the possibilities of simultaneous countdowns and collisions. However, NS2 simulation results indicate that our simple technique is already good enough for many topologies (see Section 2).

### 3.5 Accuracy of BoE versus ICN

In the above, we have explained BoE in terms of ICN with $c \to 0$. One may, therefore, expect that for nonzero $c$, ICN is more accurate than BoE. It turns out that this is not the case, at least for networks of up to 50 links. As an example, consider 802.11b when the physical-layer data rate is 11 Mbps, and the

TABLE 1
Mean Link Throughput Errors Computed Using BoE and ICN for Networks of Fixed Link Density

| # Links | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| BoE | 2.19% | 2.83% | 3.03% | 2.33% | 5.08% |
| ICN | 6.14% | 3.43% | 4.30% | 3.78% | 8.34% |
| # Links | 60 | 70 | 80 | 90 | 100 |
| BoE | 6.20% | 5.11% | 10.62% | 9.06% | 9.49% |
| ICN | 6.72% | 9.17% | 7.11% | 13.40% | 11.15% |

packet payload is 1,460 Bytes. The corresponding $c$ is 0.1867, not very close to zero. We show three examples below where BoE is more accurate than ICN. The first example is the network in Fig. 1. The simulated normalized throughputs are (0.99, 0, 0.50, 0.51), which validate the normalized throughputs computed by BoE, (1, 0, 0.5, 0.5). The normalized throughputs of ICN are (0.93, 0.08, 0.51, 0.51), slightly less accurate than BoE.

The second example is a 25-link network whose contention graph is a $5 \times 5$ grid. Let $S_{ij}$ denote the state of vertex $(i, j)$, $1 \leq i \leq 5, 1 \leq j \leq 5$. There is only one MIS in this topology in which $S_{ij} = 1$ if $(i + j)$ is even, and $S_{ij} = 0$ if $(i + j)$ is odd. According to BoE, these are the normalized throughputs of the corresponding links. NS2 simulation results match the BoE prediction. By contrast, the normalized throughputs computed from ICN are not as accurate. They range from 0.70 to 0.75 for links of even $(i + j)$, and from 0.20 to 0.22 for links of odd $(i + j)$.

For the third example, we look at randomly generated networks. For the 50-link network considered in Section 2, recall that the mean link throughput error of BoE is 5 percent. We find that mean link throughput error of ICN is higher at more than 8 percent. More generally, Table 1 shows the results for networks of 10-100 links. The network area is allowed to vary so that the link density remains the same for the networks of different sizes. On average, each link has around four neighbors. This is to simulate real-life 802.11 networks within a building in which one may see, say, up to three neighboring WLANs on the same frequency channels; we assume that link aggregation (see Section 2) can be used to aggregate the links within a WLAN into one to two virtual links. As shown in Table 1, for networks of up to 50 (virtual) links, the error of BoE is kept to 5 percent or below, while the error of ICN is consistently higher than that of BoE. For networks of 60-100 links, sometimes, BoE is more accurate than ICN, and sometimes, the other way round. The maximum error of BoE is slightly more than 10 percent.

Table 2 shows the scenario in which the network area is fixed, while the number of links is varied. That is, the link density is lower for smaller networks. Again, BoE is consistently more accurate than ICN for networks smaller than 50 links, and the error of BoE is kept below 5 percent. Our other simulation results (not shown here to conserve space) indicate that BoE can predict well for networks of up to 100 links of small link density (two links per $1{,}000\ \mathrm{m} \times 1{,}000\ \mathrm{m}$).

We noticed in our simulation results that BoE generally yields more polarized throughput distributions among links than ICN do, and the polarized throughput distributions are closer to simulation results. Further investigation indicates that there are two factors contributing to the less polarized throughput distribution observed in ICN than actuality. The first factor is that packet collisions and the resulting doubling of countdown time [18] are ignored in ICN. It turns out that the disadvantaged links may suffer more collisions than the links in the MIS. The second factor is that the operation of Extended Interframe Space (EIFS) [18] in 802.11 is ignored in ICN. The duration of EIFS is SIFS+ACK+DIFS. If a station receives an erroneous MAC frame, it will wait for an idle period of EIFS before continuing its backoff countdown. This is because the MAC frame could be targeted for another station and received correctly there, and the EIFS allows enough time for this other station to receive its ACK. It turns out that the disadvantaged links may be affected more detrimentally by the EIFS operation. These two factors cause the "effective $c$" of the disadvantaged links to be larger than that of the links in the MIS, since the countdown overhead then becomes larger.

TABLE 2
Mean Link Throughput Errors Computed Using BoE and ICN for Networks of Fixed Area ($3{,}160\ \mathrm{m} \times 3{,}160\ \mathrm{m}$)

| # Links | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| BoE | 1.41% | 1.38% | 3.96% | 4.87% | 5.08% |
| ICN | 6.06% | 4.49% | 6.40% | 9.18% | 8.34% |

For verification, we disabled doubling of contention window CW upon collisions, set $\mathrm{EIFS} = \mathrm{DIFS}$, and performed the NS2 simulation again. This time, the NS2 throughput distribution is the same as the ICN throughput distribution. Of course, the EIFS and doubling of CW upon collisions in 802.11 are there for good reason, and we should not deactivate them in actual network operation. To make ICN more accurate, one could augment it with a "perturbation analysis" to take into account the two factors. For modest-size networks, however, we find that the simple method implied by BoE that takes $c \to 0$ in ICN is already highly accurate.

While BoE is very good for networks with up to 50 links, it may become less accurate for very large networks. BoE neglects all the non-MIS states in a network. This is reasonable in a modest-size network. As network size increases, however, the number of IS that is not MIS also increases relative to the number of MIS. Even though each of the non-MIS may have a small probability, there could be a huge number of such states. As a result, the sum of their probabilities cannot be ignored any more. Indeed, our experiments above showed that for larger random networks, sometimes, ICN (which does not ignore non-MIS) gives better predictions.

## 4 Insensitivity to Backoff and Transmission Time Distributions

In this section, we prove that $P_s$ as expressed in (4) is insensitive to the distributions of backoff countdown and transmission times, $f(t_{cd})$ and $g(t_{tr})$, given the ratio of their means $c$. This result is important because for practical

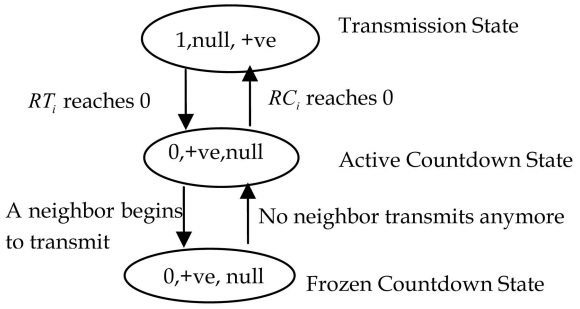



Fig. 6. The state-transition diagram of an ideal link.

CSMA protocols (e.g., 802.11) and network applications (e.g., FTP and P2P file download), neither the countdown nor transmission time is exponentially distributed.

### 4.1 General ICN Model

Previously, we modeled the link state in ICN with a binary variable $S_i$, where $S_i = 0$ if link $i$ is not transmitting and $S_i = 1$ if link $i$ is transmitting. This model is complete under the exponential countdown and transmission assumption because of the memoryless property of the exponential distribution.

To model link state in general, we need to provide more details to capture the remaining countdown time when $S_i = 0$; and the remaining transmission time when $S_i = 1$. Specifically, we define the detailed state of a link $i$ as $X_i = (S_i, RC_i, RT_i)$, where $RC_i, RT_i \in [0, +\infty)$ are the residual countdown and transmission times, respectively. Since there are infinite possible values for $RC_i$ and $RT_i$, the state space for the detailed state is infinite. The infinite state space can, in turn, be partitioned into three coarse states, as follows:

**Transmission state.** Link $i$ is in the midst of transmission. The possible detailed-state values for $(S_i, RC_i, RT_i) = (1, \text{ null}, +ve)$. That is, when $S_i = 1$, $RT_i$ is a positive real number, and the value of $RC_i$ is irrelevant.

**Active countdown state.** Link $i$ is in the midst of active countdown. The possible detailed-state values are $(S_i, RC_i, RT_i) = (0, +ve, \text{ null})$.

**Frozen countdown state.** Link $i$ is in the midst of countdown, but its countdown is being frozen because at least one of its neighbors is transmitting. The possible detailed-state values are $(S_i, RC_i, RT_i) = (0, +ve, \text{null})$.

The transition diagram for the coarse state is depicted in Fig. 6. When a link begins to transmit, the transmission time $T_{tr}$ is generated according to the probability density $g(t_{tr})$. The initial residual transmission time $RT_i$ is set to the $T_{tr}$ generated, and its rate of decrease is $dRT_i/dt = -1$. When $RT_i$ reaches zero, the link enters the active countdown state. The countdown time $T_{cd}$ is generated according to the probability density $f(t_{cd})$. The initial residual countdown time $RC_i$ is set to the $T_{cd}$ generated, and its rate of decrease is $dRC_i/dt = -1$. When $RC_i$ reaches zero, the link reenters the transmission state, and a new $T_{tr}$ is generated. However, it is possible that a neighbor begins to transmit, while the link is in the midst of active countdown, in which case the link enters the frozen countdown state, during which $dRC_i/dt = 0$. It is possible that other neighbors also begin to transmit, while the link is in the frozen countdown state. When the transmissions of all neighbors have ended and no neighbor is left transmitting, the link begins active countdown again, continuing with the $RC_i$ value when it was last frozen (i.e., countdown has memory).

Note that there is no direct transition between transmission and frozen countdown states because of the assumption of continuous countdown time. For the link to go directly from transmission state to frozen state, the link has to complete its transmission exactly at the same time that a neighbor completes its countdown; but the probability of this event is zero under the continuous $T_{cd}$ assumption.

### 4.2 Insensitivity Proof

Our technical report [8] presents three insensitivity proofs yielding different insights to the characteristics of ICN (e.g., time reversibility of $S = S_1 \dots S_L$, etc.). Interested readers are referred to [8] for details. We choose to present here the shortest proof among the three proofs.

Let $X = X_1 X_2 \cdots X_L$ be the overall detailed system state, and $x = x_1 x_2 \cdots x_L$ be a particular realization of $X$. In addition, let $s = s_1 s_2 \cdots s_L$ be the realization of $S = S_1 S_2 \cdots S_L$ under $x$. Similarly, let $rc = rc_1 rc_2 \cdots rc_L$ and $rt = rt_1 rt_2 \cdots rt_L$ be the realizations of residual countdown and transmission times, $RC = RC_1 RC_2 \cdots RC_L$, and $RT = RT_1 RT_2 \cdots RT_L$ , respectively, under $x$. Let $\chi$ be the set of all feasible states $x$. We note that $X$ is a **continuous-state Markov process**. The balance equation for the continuous-state Markov process is in the form of a partial differential equation derived below.

**Derivation of Balance Equation.**

Let us first derive the transient balance equation. Let $p_X(t, x)$ be the state probability density at time $t$. We want to derive an expression for

$$\frac{dp_X(t,x)}{dt} = \lim_{\Delta t \to 0} \frac{p_X(t + \Delta t, x) - p_X(t, x)}{\Delta t}. \tag{7}$$

In the time interval from $t$ to $t + \Delta t$, the state changes as a result of links counting down and transmitting. There are two types of "jump events" that cause a discontinuity in the evolution of $x$: 1) a link finishing a transmission and 2) a link finishing counting down. For small $\Delta t$, the probability of having more than one jump event in the overall system is of order $o(\Delta t)$. Between the jump events, the links that are transmitting continue to transmit and the links that are counting down continue to count down, with $drt_i/dt = -1$ and $drc_i/dt = -1$, respectively, in which case $x$ changes continuously without $s$ being changed. For a particular state realization $x$, we can write

$$p_X(t + \Delta t, x) = \text{RT0} + \text{RC0} + \text{CDTR} + o(\Delta t), \tag{8}$$

where RT0 is the contribution due to end-of-transmission jump events, RC0 is the contribution due to countdown-to-zero jump events, CDTR is the contribution due to ordinary counting down and transmission without any jump events, and $\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$.

Let $G$ denote the set of all links in the network. Under a specific realization $x$, let $T(x)$ be the set of transmitting links, and $C(x)$ be the set of links that are actively counting

down. The links that are in the frozen countdown state are, therefore, $G - T(x) - C(x)$. We can write

$$\begin{aligned}\text{RT0} &= \sum_{i\in C(x)} p_X(t, S_{i1}RT_{i0^+}x)\cdot \Delta rt_i \cdot f(rc_i)\\ &= \sum_{i\in C(x)} p_X(t, S_{i1}RT_{i0^+}x) f(rc_i)\Delta t,\end{aligned} \tag{9}$$

where $RT_{i0^+}$ is the operator that sets $rt_i$ in $x$ to $0^+$ (i.e., just before transmission of link i completes), $S_{i1}$ is the operator that sets $s_i$ in $x$ to 1, and $f(rc_i)$ is the probability density of a newly generated countdown time. Note that in the above, for link $i$ to just finish transmission within $\Delta t$, $rt_i$ must fall within the interval $(0, \Delta rt_i)$ at time $t$, where $\Delta rt_i = |drt_i/dt|\Delta t = \Delta t$. Similarly, we have

$$\begin{aligned}\text{RC0} &= \sum_{i\in T(x)} p_X(t, S_{i0}RC_{i0^+}x)\cdot \Delta rc_i \cdot g(rt_i)\\ &= \sum_{i\in T(x)} p_X(t, S_{i0}RC_{i0^+}x) g(rt_i)\Delta t,\end{aligned} \tag{10}$$

where $RC_{i0^+}$ is the operator that sets $rc_i$ in $x$ to $0^+$ (i.e., just before countdown of link $i$ completes), $S_{i0}$ is the operator that sets $s_i$ in $x$ to 0, and $g(rt_i)$ is the probability density of a newly generated transmission time.

For CDTR, in order to evolve to state $x$ at time $t + \Delta t$, each link $i$ that is actively counting down must have $RC_i = rc_i + \Delta t$, and each link that is transmitting must have $RT_i = rt_i + \Delta t$ at time $t$. That is, $\Delta rc_i = \Delta t$ and $\Delta rt_i = \Delta t$, respectively. By Taylor expansion, we have

$$\begin{aligned}\text{CDTR} &= p_X(t, x + \Delta x) = p_X(t, x)\\ &\quad + \sum_{i\in C(x)} \frac{\partial p_X(t,x)}{\partial rc_i}\Delta t + \sum_{i\in T(x)} \frac{\partial p_X(t,x)}{\partial rt_i}\Delta t + o(\Delta t).\end{aligned} \tag{11}$$

Putting (9), (10), and (11) into (8), and then, taking the derivative limit in (7), we get

$$\begin{aligned}\frac{dp_X(t,x)}{dt} &= \sum_{i\in C(x)} \frac{\partial p_X(t,x)}{\partial rc_i} + \sum_{i\in T(x)} \frac{\partial p_X(t,x)}{\partial rt_i}\\ &+ \sum_{i\in C(x)} p_X(t, S_{i1}RT_{i0^+}x) f(rc_i) + \sum_{i\in T(x)} p_X(t, S_{i0}RC_{i0^+}x) g(rt_i).\end{aligned} \tag{12}$$

The equilibrium balance equation (13) below is obtained from the transient balance equation (12) by noting that at equilibrium, $dp_X(t,x)/dt = 0$ and $p_X(t,x)$ approaches the steady-state probability density $p_X(x) \forall x \in \chi$:

$$\begin{aligned}&-\sum_{i\in C(x)} \frac{\partial p_X(x)}{\partial rc_i} - \sum_{i\in T(x)} \frac{\partial p_X(x)}{\partial rt_i}\\ &= \sum_{i\in C(x)} p_X(S_{i1}RT_{i0^+}x) f(rc_i) + \sum_{i\in T(x)} p_X(S_{i0}RC_{i0^+}x) g(rt_i).\end{aligned} \tag{13}$$

The remaining step is to find the solution to (13). Consider a network with one and only one isolated link $i$, alternating between counting down and transmission. If we choose random points to observe $RC_i$ while the link is counting down, it is straightforward to show that the probability density of $RC_i$ is related to the probability distribution of countdown time by $f_{RC_i}(rc_i) = (1 - F(rc_i))/E[T_{cd}]$, where $F(rc_i) = \int_0^{rc_i} f(t_{cd})dt_{cd}$. Similarly, random observations of $RT_i$ while the link is transmitting yield probability density of $RT_i$ given by $g_{RT_i}(rt_i) = (1 - G(rt_i))/E[T_{tr}]$, where $G(rt_i) = \int_0^{rt_i} g(t_{tr})dt_{tr}$. It turns out that these distributions of residual countdown and transmission times remain valid under random observations of a general CSMA network with many interacting links rather than just one isolated link. This fact is embodied in Theorem 1 below.

**Theorem 1.** *The equilibrium probability density of X is*

$$p_X(x) = P_s \prod_{i\in T(x)} g_{RT_i}(rt_i) \prod_{i\in G-T(x)} f_{RC_i}(rc_i) \quad \forall x \in \chi, \tag{14}$$

*where $P_s$ is given by (4).*

**Comment.** By integrating $p_X(x)$ in (14) over all possible values of $rt_i$ and $rc_i$ for all $i$, we get $P_s$ in (4). In other words, the distribution of $S$ is insensitive to the forms of $f(t_{cd})$ and $g(t_{cd})$.

**Proof.** We show that (14) satisfies the balance equation (13). In fact, the terms in the LHS and RHS of (13) match on a one-to-one basis under (14). In the following, we show that $-\frac{\partial p_X(x)}{\partial rc_j} = p_X(S_{j1}RT_{j0^+}x) f(rc_j)$ for each $j \in C(x)$ under (14). The argument for $-\frac{\partial p_X(x)}{\partial rt_j} = p_X(S_{j0}RC_{j0^+}x) g(rt_j)$ for each $j \in T(x)$ is similar. Under (14),

$$\begin{aligned}-\frac{\partial p_X(x)}{\partial rc_j} &= -\frac{df_{RC_j}(rc_j)}{drc_j} P_s \prod_{i\in T(x)} g_{RT_i}(rt_i) \prod_{i\in G-T(x)-j} f_{RC_i}(rc_i)\\ &= \frac{f(rc_j)}{E[T_{cd}]} \frac{B}{c^{|T(x)|}} \prod_{i\in T(x)} g_{RT_i}(rt_i) \prod_{i\in G-T(x)-j} f_{RC_i}(rc_i).\end{aligned} \tag{15}$$

Under (14), we also have

$$\begin{aligned}&p_X(S_{j1}RT_{j0^+}x) f(rc_j)\\ &= f(rc_j) \frac{B\cdot g_{RT_j}(0^+)}{c^{|T(x)|+1}} \prod_{i\in T(x)} g_{RT_i}(rt_i) \prod_{i\in G-T(x)-j} f_{RC_i}(rc_i)\\ &= \frac{f(rc_j)}{E[T_{tr}]} \frac{B}{c^{|T(x)|+1}} \prod_{i\in T(x)} g_{RT_i}(rt_i) \prod_{i\in G-T(x)-j} f_{RC_i}(rc_i).\end{aligned} \tag{16}$$

Thus, (15) and (16) are equal. □

In MATLAB simulations, perfect fits with (14) are obtained for various distributions, including uniform distribution for countdown time and fixed transmission time (where $g(\cdot)$ is a $\delta(\cdot)$ function). The link throughputs and airtimes occupied by are also exactly as predicted by (4).

## 5 Implications, Applications, and Extensions

This section is devoted to the discussion of several interesting implications arising out of the studies of BoE and ICN. In addition, we also present several applications and new research directions stemming from BoE and ICN.

### 5.1 Global Optimality and Local Starvation/ Unfairness

First of all, recall from discussion in Section 3.5 that $c$ does not have to be extremely small for BoE approximation to be good for 802.11 networks of up to 50 links. That is, even for moderately small $c$, after taking into other aspects of 802.11, the system spends almost all its time among the MIS. This implies that **the highest global throughput** is achieved, since the number of simultaneously transmitting links is the highest possible in MIS. On the other hand, starvation of some links is a common phenomenon (e.g., Fig. 2). Indeed, an application of BoE is to quickly identify starved links so that remedial actions could be taken. The remedy could be to assign the starved links to other frequency channels, or to vary $c$ among the links.

For the latter, we will need to move away from BoE (which ignores non-MIS) to consider the general ICN (which includes non-MIS as part of the overall analysis). Consider the network of Fig. 1 in which link 2 is starved. To unstarve it, one could change its $c$, either by reducing the average countdown time, or increasing the average transmission duration (i.e., CW and TXOP, respectively, in the parlance of 802.11 [19]). Reducing $c_2$ causes the non-MIS state 0100 in ICN to have non-negligible probability, and detailed-balance analysis similar to that of (4) (see (17)) can be used to set $c_2$. The reader could verify that with $c_1 = 1.1111, c_2 = 0.0584, c_3 = c_4 = 0.1111$, the links would have equal throughput of 0.32.

### 5.2 Resource Allocation

From a "resource allocation" standpoint, the CSMA protocol is just **a distributed implementation** of resource allocation with an **implicit utility objective**. BoE is not a method for resource allocation, but a computation method that computes the resulting throughput distribution as dictated by this implicit utility objective. For typical default parameter settings of 802.11 networks, BoE states that only the MISs among all the IS are allocated appreciable airtime, and all the MISs are allocated equal airtime. The implicit **utility** being maximized is the **global system throughput**.

More generally, the utility to be optimized in ICN can be modified by adjusting $c_i, i = 1, \ldots, L$, of different links. The generalization of (4) under ICN is $P_s = B/\prod_{i|s_i=1 \text{ in } s} c_i$. The resource allocation problem as a mathematical optimization problem can be formulated as

$$\begin{aligned} &\text{Max } U(Th_1, Th_2, \ldots, Th_L) \\ &\text{s.t. } P_s = \frac{B}{\prod_{i|s_i=1 \text{ in } s} c_i} \forall\ s, \\ &\qquad Th_i = \sum\nolimits_{s:s_i=1} P_s \ \forall\ i, \end{aligned} \tag{17}$$

where $U(Th_1, Th_2, \ldots, Th_L)$ is the utility as a function of individual link throughputs. A possible utility function, for example, could be the proportional fairness objective [20]. Optimization problems such as (17) have been well studied in the context of time-division multiple-access (TDMA) networks. Rather than the constraints in TDMA networks, the dynamic of CSMA networks results in constraints in the form as shown in (17).

An insight of this paper is that under ICN, the lengths and distributions of countdown and transmission times are immaterial and cannot be used for resource allocation purposes once the ratio of their means is given. We only have $L$ degrees of freedom for our resource allocation problem: $c_i, i = 1, \ldots, L$.[5]

### 5.3 Small Change in Topology Leads to Large Change in Throughputs of Faraway Links

According to BoE, it may be difficult to estimate the throughput of a link based on the "local" topology around it. Another link at a distance can affect it significantly. Take topology 5 in Fig. 2. All five links in the ring have the same throughput. The addition of link 6 in topology 6, however, causes three links to be starved and three links to grab the maximum throughput. This observation can be extrapolated to a large ring consisting of $L \gg 1$ vertices. For odd $L$, there are $L$ MIS and each vertex is in $(L-1)/2$ of the MIS. The throughput of each link approaches 0.5 for large $L$. However, the addition of an extra link as in topology 6 causes a drastic in throughput distribution, even for links that are very far away from the location of the perturbation. This is a consequence of letting $c \to 0$ in our ICN.

The above observation, when mapped to practice in 802.11 networks, raises an interesting issue. To reduce overhead, it is desirable to increase the TXOP [19] so that whenever a link grabs the channel, it sends as much data as possible. The consequence is a small $c$. A smaller $c$, however, has the above coupling effects among links that are far apart, making distributed network design more challenging.

### 5.4 Island States

The throughputs computed by BoE and ICN are the long-term averages obtained from stationary distributions. Temporal starvation may still occur despite acceptable long-term averages because of the difficulty of movement between states. To see this, consider topology 7 in Fig. 2. The MISs are $\begin{smallmatrix}1&0&1\\0&1&0\end{smallmatrix}$ and $\begin{smallmatrix}0&1&0\\1&0&1\end{smallmatrix}$. According to BoE, the long-term throughput of each link is 0.5. However, the Hamming distance between these two MIS is six, meaning that the states of six links have to change in order to move from one MIS to the other MIS. When $c$ is small and the Hamming distance between two MIS is large, such a move occurs only rarely because the small $c$ tends to "push" the system back to an MIS whenever it tries to move away in its random walk on the state-transition diagram. It is as if the two MISs are islands separated by oceans. **Temporal equilibrium** stays around an MIS with only occasional movement across the two MISs. In topology 7, three links can starve for a long time once the system settles around the MIS that disfavors them.

Temporal starvation and the existence of island states can be identified from the state-transition diagram and Hamming distance analysis. Such "long-term oscillatory behavior" has also been observed in [22] in its TCP over 802.11 wireless network simulation results. Our work here suggests that the root cause may not be TCP per se, but that CSMA networks can settle into different temporal equilibriums under greedy competition.

5. Subsequent to the publication of our work and the description of this resource allocation problem in [8], Jiang and Walrand [21] succeeded in finding an elegant distributed algorithm to (17) without message passing between the links.

### 5.5 Unsaturated Networks

This paper has focused on networks with saturated links competing in a greedy manner (which well models the behavior of links that always have traffic to send—e.g., TCP traffic arising from long-lasting P2P file downloads). For networks with unsaturated links, there are two possible extremes. The first is the bursty traffic case in which each link becomes saturated and idle alternatively, with the saturated and idle periods each lasting a long time. The second is the intermittent traffic case in which the input traffic of links (i.e., their offered load) is below their saturated throughputs, but the traffic arrives to the links continually.

The first case can be viewed as one in which the network topology changes dynamically dynamic. At any instance, the network consists only of those links with saturated traffic. As time progresses, there is a sequence of effective network topologies. The BoE method (or ICN method) can be used to compute the throughput distributions under each of the network topologies.

The second case is more challenging, particularly if we are interested in not just the access delay, but also the overall queuing delay which includes the buffer delay at a link. This is an interesting subject for further investigation.

### 5.6 Networks with Hidden Nodes

The BoE method and the ICN model presented in this paper assume that there are no hidden nodes [23] in the CSMA network. A network without hidden nodes can be designed based on the principle of hidden-node free design (HFD) described in [23]. BoE and ICN are compatible with networks designed according to the HFD principle. Extensions to the case in which there are hidden nodes is an interesting subject for further studies.

### 5.7 Relationship with Statistical Physics

It turns out the ICN is similar to the **Ising model** in statistical physics in which each entity can take on one of two possible states. Statistical physics deals with statistical manifestations of a large number of interacting entities. It is, therefore, not surprising that when $L$ is large, the problem domain of the CSMA network becomes similar to that of statistical physics.

The results from the study of Ising model may shed light on the dynamic in CSMA networks. In the Appendix, we show that $P_s$ is an **MRF** [8], [11]. MRF was originally motivated from statistical physics to capture the "spatial" interdependencies of stochastic processes. It will be interesting to explore whether the MRF formulation will yield further insights into CSMA networks.

## 6 Conclusions

We have presented a simple BoE method for computing throughput distributions among links in CSMA networks. For 802.11 networks with up to 50 links, BoE has been verified to be very accurate (link throughput estimation error of 5 percent or below), for both UDP and TCP traffic. The technique of link aggregation can be used in BoE to deal with even larger networks in which many links have the same carrier-sensing relationships with other links. This is typically the case, for example, within a building with multiple coexisting 802.11 WLANs.

An insight arising out of BoE is as follows: In an 802.11 network with $L$ links competing in a greedy manner, there is a maximum of $2^L$ possible states in terms of who is transmitting and who is not. However, only very few of these states are probable, and the probable states are equally probable. An immediate application of BoE is for quick identification of starved links in the network so that remedies can be devised to solve the problem.

BoE was first discovered from experimental observation rather than from analytical construction. We subsequently developed a theory to explain BoE based on an ICN model. We find that the throughputs of links in ICN are insensitive to the distributions of the backoff time and transmission time given the ratio of their mean $c$. The BoE computation method emerges from ICN in the limit $c \to 0$ (in practice, we find that $c$ does not have to be very small for BoE to be highly accurate, e.g., $c = 0.1867$ for 802.11b). The insensitivity result explains why BoE works so well for 802.11 networks, in which the backoff countdown process is one that has memory and the transmission time can be arbitrarily distributed. This is the first paper to prove the insensitivity result mathematically.

The domain of application of BoE is mainly for modest-size networks, such as the infrastructure WLANs typically seen within a building today. For very large networks, the assumption of $c \to 0$ implicit in BoE may cause non-negligible errors in the computed results. Furthermore, the computation may become intractable for both BoE and ICN. Our ongoing research focuses on quick and accurate approximate methods for very large networks based on the groundwork established in this paper.

## Appendix
## $P_s$ is a Markov Random Field

Consider a system (not necessarily ICN) consisting of $L$ entities. The relationships between the entities are modeled by a graph $G$ in which vertex $i$ corresponds to entity $i$, and an edge joins two vertices if they could interact with each other. The value of the system state is a vector consisting of the values of individual states of the entities, $s = s_1 s_2 \ldots s_L$, with stationary probability distribution $P_s$. $P_s$ is said to be a Markov random field if $P_{s_i|s_{G-i}} = P_{s_i|s_{N_i}}$, where $s_{G-i}$ denotes the states of all other entities in the graph $G$ except $s_i$, $N_i$ denotes the neighbors of $i$, and $s_{N_i}$ denotes the states of the neighbors of $i$ only. That is, given the states of its neighbors, the state of entity $i$ is independent of the states of all other entities in the system.

It turns out that $P_s$ in ICN is a Markov random field: the "only if" part of the proof of Theorem 2 below shows that the probability distribution given by (4) is a Markov random field. The "if" part of the proof of Theorem 2 can be considered as an alternative proof of the insensitivity of (4) to $f(t_{cd})$ and $g(t_{cd})$, starting with the premise that $P_s$ is a Markov random field.

**Theorem 2.** *$P_s$ of ICN is given by (4) if and only if $P_{s_i|s_{G-i}} = P_{s_i|s_{N_i}}$ for all $i$ .*

**Proof.** If—Consider an arbitrary link $i$. Let $P_{s_i,s_{G-i}}$ denote the probability of the system state in which the state of link $i$ adopts the value of $s_i$, and the states of other links adopt the values in $s_{G-i}$. By the nature of ICN, over a long stretch

of time, after excluding the airtime during which link $i$ is in frozen countdown state, the ratio of the airtimes used for active countdown and transmission for link $i$ must be $c$. Thus, we have $\sum_{s_{G-i}} P_{0,s_{G-i}} = c \sum_{s_{G-i}} P_{1,s_{G-i}}$, where the summation is over all $s_{G-i}$ such that the system states $0, s_{G-i}$ and $1, s_{G-i}$ are connected. Note that for each of such $s_{G-i}$, link $i$ is actively counting down in state $0, s_{G-i}$ and transmitting in state $1, s_{G-i}$. Therefore, in each of the $s_{G-i}$, $s_j = 0$ for all $j \in N_i$ because link $i$ can actively count down or transmit only if its neighbors are not transmitting. Thus, we can write

$$\sum_{s_{G-i-N_i}} P_{0,\underline{0},s_{G-i-N_i}} = c \sum_{s_{G-i-N_i}} P_{1,\underline{0},s_{G-i-N_i}}, \tag{A1}$$

where the $\underline{0}$ in the indices of $P_{0,\underline{0},s_{G-i-N_i}}$ and $P_{1,\underline{0},s_{G-i-N_i}}$ is an indication that $s_j = 0$ for all $j \in N_i$; and $s_{G-i-N_i}$ are the states of links who are not neighbors of $i$. For a pair of connected states $0, s_{G-i}$ and $1, s_{G-i}$, we can write

$$\frac{P_{0,s_{G-i}}}{P_{1,s_{G-i}}} = \frac{P_{0,\underline{0},s_{G-i-N_i}}}{P_{1,\underline{0},s_{G-i-N_i}}} = \frac{P_{0|\underline{0},s_{G-i-N_i}}}{P_{1|\underline{0},s_{G-i-N_i}}} = \frac{P_{0|\underline{0}}}{P_{1|\underline{0}}}, \tag{A2}$$

where the $\underline{0}$ in the indices $0|\underline{0}$ and $1|\underline{0}$ denotes the event that $s_j = 0$ for all $j \in N_i$; the first equality is due to the CSMA operation; and last equality is due to the Markov-random field property $P_{s_i|s_{G-i}} = P_{s_i|s_{N_i}}$. Plugging (A2) into (A1) gives $P_{0|\underline{0}}/P_{1|\underline{0}} = c$, and replugging this back into (A2) yields the detailed balance equation that leads to (4).

Only if—For the case where $s_j = 1$ for some $j \in N_i$, by the nature of the system, it is clear that $P_{s_i|s_{G-i}} = P_{s_i|s_{N_i}}$ because the conditional probability is zero if $s_i = 1$, and is one if $s_i = 0$ regardless of $s_{G-i-N_i}$. For the case where $s_j = 0$ for all $j \in N_i$, detailed balance gives

$$\begin{aligned}
&P_{0,\underline{0},s_{G-i-N_i}} = cP_{1,\underline{0},s_{G-i-N_i}} \forall s_{G-i-N_i} \\
&\Rightarrow P_{0|\underline{0},s_{G-i-N_i}} = cP_{1|\underline{0},s_{G-i-N_i}} \forall s_{G-i-N_i} \\
&\Rightarrow P_{0|\underline{0},s_{G-i-N_i}} = c/(1+c), P_{1|\underline{0},s_{G-i-N_i}} = 1/(1+c) \forall s_{G-i-N_i}.
\end{aligned}$$

Thus, $P_{s_i|\underline{0},s_{G-i-N_i}}$ is independent of $s_{G-i-N_i}$ and $P_{s_i|s_{G-i}} = P_{s_i|s_{N_i}}$. □

## ACKNOWLEDGMENTS

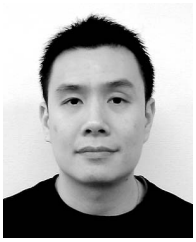
This work was supported by the Competitive Earmarked Research Grant (Project Number 414507) established under the University Grant Committee of the Hong Kong Special Administrative Region, China, and the Direct Grant (Project Number 2050436) from the Chinese University of Hong Kong.

## REFERENCES

[1] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Selected Areas in Comm.*, vol. 18, no. 3, pp. 535-547, Mar. 2000.
[2] P.C. Ng and S.C. Liew, "Throughput Analysis of IEEE802.11 Multi-Hop Ad-Hoc Networks," *IEEE/ACM Trans. Networking*, vol. 15, no. 2, pp. 309-322, Apr. 2007.
[3] P.C. Ng and S.C. Liew, "Offered Load Control in IEEE 802.11 Multi-Hop Ad-Hoc Networks," *Proc. IEEE Int'l Conf. Mobile Ad-Hoc and Sensor Systems (MASS)*, 2004.
[4] G. Yan, D.-M. Chiu, and J.C.S. Lui, "Determining the End-to-End Throughput Capacity in Multi-Hop Networks: Methodology and Applications," *Proc. ACM Sigmetric*, 2006.
[5] M. Garetto, T. Salonidis, and E.W. Knightly, "Modeling Per-Flow Throughput and Capturing Starvation in CSMA Multi-Hop Wireless Networks," *Proc. IEEE INFOCOM*, 2006.
[6] K. Wang, F. Yang, Q. Zhang, and Y. Xu, "Modeling Path Capacity in Multi-Hop IEEE 802.11 Networks for QoS Services," *IEEE Trans. Wireless Comm.*, vol. 6, no. 2, pp. 738-749, Feb. 2007.
[7] K. Medepalli and F.A. Tobagi, "Towards Performance Modeling of IEEE 802.11 Based Wireless Networks: A Unified Framework and Its Applications," *Proc. IEEE INFOCOM*, 2006.
[8] S.C. Liew, C. Kai, B. Wong, and J. Leung, "Back-of-the-Envelope Computation of Throughput Distributions in CSMA Wireless Networks," Technical Report, http://arxiv.org/pdf/0712.1854.pdf, Dec. 2007.
[9] R.R. Boorstyn et al., "Throughput Analysis in Multihop CSMA Packet Radio Networks," *IEEE Trans. Comm.*, vol. 35, no. 3, pp. 267-274, Mar. 1987.
[10] X. Wang and K. Kar, "Throughput Modelling and Fairness Issues in CSMA/CA Based Ad-Hoc Networks," *Proc. IEEE INFOCOM*, 2005.
[11] M. Durvy and P. Thiran, "A Packing Approach to Compare Slotted and Non-Slotted Medium Access Control," *Proc. IEEE INFOCOM*, 2006.
[12] F.P. Kelly, "Loss Networks," *Annals of Applied Probability*, vol. 1, no. 3, pp. 319-378, 1991.
[13] D.Y. Burman, J.P. Lehoczky, and T. Lim, "Insensitivity of Blocking Probabilities of a Circuit-Switched Network," *J. Applied Probability*, vol. 21, no. 4, pp. 850-859, 1984.
[14] "MRF," http://en.wikipedia.org/wiki/Markov_network, 2010.
[15] F.P. Kelly, *Reversibility and Stochastic Networks*. Wiley, 1979.
[16] "Maximum Independent Set," http://en.wikipedia.org/wiki/Maximal_independent_set, 2010.
[17] "MADwifi," http://madwifi.org/wiki, 2008.
[18] *IEEE Std 802.11-1997, IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE, 1997.
[19] *IEEE Std 802.11e-2005, IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Quality of Service Enhancements*, IEEE, 2005.
[20] S.C. Liew and Y.J. Zhang, "Proportional Fairness in Multi-Channel Multi-Rate Wireless Networks—Part I: The Case of Deterministic Channel; Part II: The Case of Time-Varying Channel," *IEEE Trans. Wireless Comm.*, vol. 7, no. 9, pp. 3446-3456 and 3457-3467, Sept. 2008.
[21] L. Jiang and J. Walrand, "A Distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks," *IEEE/ACM Trans. Networking*, vol. 18, no. 3, pp. 960-972, June 2010.
[22] V. Raghunathan and P.R. Kumar, "A Counterexample in Congestion Control of Wireless Networks," *Performance Evaluation*, vol. 64, no. 5, pp. 399-418, June 2007.
[23] L.B. Jiang and S.C. Liew, "Hidden-Node Removal and Its Application in Cellular WiFi Networks," *IEEE Trans. Vehicular Technology*, vol. 56, no. 5, pp. 2641-2654, Sept. 2007.

**Soung Chang Liew** received the SB, SM, EE, and PhD degrees from the Massachusetts Institute of Technology. From March 1988 to July 1993, he was at Bellcore (now Telcordia), New Jersey, where he was engaged in Broadband Network Research. He has been a professor in the Department of Information Engineering, The Chinese University of Hong Kong, since 1993. His current research interests focus on wireless networking. He and his student won the best paper awards at the First IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (IEEE MASS 2004) and the Fourth IEEE International Workshop on Wireless Local Network (IEEE WLN 2004). Separately, TCP Veno, a version of TCP to improve its performance over wireless networks proposed by him and his student, has been incorporated into a recent release of Linux OS. In addition, he initiated and built the first interuniversity ATM network testbed in Hong Kong in 1993. More recently, his research group pioneers the concept of Physical-layer Network Coding (PNC) for application in wireless networks. Besides academic activities, he cofounded two technology start-ups in Internet Software and has been serving as a consultant to many companies and industrial organizations. He is currently a consultant for the Hong Kong Applied Science and Technology Research Institute (ASTRI), providing technical advice as well as helping to formulate R&D directions and strategies in the areas of Wireless Internetworking, Applications, and Services. He is the holder of six US patents and a fellow of the IEE and HKIE. He is listed in *Marquis Who's Who in Science and Engineering*. He is the recipient of the first Vice-Chancellor Exemplary Teaching Award at The Chinese University of Hong Kong. He is a senior member of the IEEE and the IEEE Computer Society. More details about his research can be found at www.ie.cuhk.edu.hk/soung.

**Cai Hong Kai** received the BE degree from Hefei University of Technology in 2003, and the ME degree in electronic engineering and computer science from the University of Science and Technology of China in 2006. She is currently working toward the PhD degree at the Department of Information Engineering, The Chinese University of Hong Kong. Her research interests lie in the areas of wireless networking, including design and analysis of medium access control protocols for wireless networks, TCP improvement schemes in wireless networks, and cross-layer design.



**Hang Ching (Jason) Leung** received the graduate degree and the MPhil degree in electronic and electrical engineering from the Hong Kong University of Science and Technologies. He is a senior engineer at Altai Technologies. He leads the development of advance technologies in wireless network systems. His study was focused on baseband design of advance coding schemes in wireless MIMO systems. After graduation, he joined Hong Kong Applied Science & Technologies Research Institute (ASTRI) and continued his work on wireless networks, focusing on system-level optimization, including load balancing, frequency assignment, and many other key issues in wireless network management. Later on, he brought his work to Altai Technologies, a spin-off from ASTRI, and participated in the development of cellular WiFi equipment, including WiFi base stations and RF controllers with world-leading coverage performance. With more than eight years of R&D experience in the wireless domain, his knowledge in data network architectures and wireless RF techniques is both wide-ranging and in depth.

**Piu (Bill) Wong** received the BS degree in electrical engineering from Rose-Hulman Institute of Technology in Indiana, and the MS and PhD degrees in electrical engineering from Stanford University. He is the chief technology officer and the vice president of engineering at Altai Technologies, responsible for defining the overall technical strategies and product development for the company. Prior to Altai Technologies, he was the founder and the vice president of the Wireless Access Group (WAG) at Hong Kong Applied Science and Technology Research Institute Company Limited (ASTRI). During his tenure, he established the WAG and delivered commercial WiFi infrastructure products with smart antenna technologies. The WAG was successfully spun off from ASTRI to form Altai Technologies, Ltd. He has more than 15 years of advanced technology and product development in wireless communication systems. He brings with him a successful track record in wireless start-up business from Silicon Valley. He cofounded Silicon Telecom, which specialized in PACS and PHS smart antenna base station equipment. He was then one of the cofounders of Adaptive Telecom, Inc. (ATI) developing smart antenna technologies for CDMA. After successfully demonstrating a smart antenna system, ATI was subsequently acquired by an NASDAQ-listed company for US $130 million with 30X return for its VC investors. He owns more than 15 US patents and with many pending. He has numerous publications, including the IEEE journals, and many invited presentations in major international conferences.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.