Improving Throughput and Fairness by Reducing Exposed and Hidden Nodes in 802.11 Networks

Li Bin Jiang, Soung Chang Liew, *Senior Member, IEEE*

Abstract—Two well known problems that can cause performance degradations in IEEE 802.11 wireless networks are the exposed-node (EN) and hidden-node (HN) problems. While there have been isolated and incidental studies of EN and HN, a comprehensive treatment has not been attempted. The contributions of this paper are three-fold. First, we provide rigorous mathematical definitions for EN and HN in wireless networks (including Wireless LANs with multiple Access Points, and ad-hoc networks). Second, we relate EN to non-scalability of network throughput; and HN to unfair throughput distributions. Third, we provide schemes to eliminate EN and HN respectively. We show that the standard 802.11 technology is not scalable because, due to EN, more Access Points (APs) do not yield higher total throughput. By removing EN, our schemes make it possible to achieve scalable throughput commensurate with the seminal theoretical results in [1], [2]. In addition, by removing HN, our schemes solve the performance problems triggered by HN, including throughput unfairness/starvation and re-routing instability.

Index Terms—IEEE 802.11, Hidden Node problem, Exposed Node Problem, Mathematical Modeling, Algorithms, Protocols, Performance Evaluation.

1 Introduction

With the increased popularity of IEEE 802.11 technology [3], it is becoming common for multiple overlapping Wireless LANs (WLANs) to be deployed over the same area. This gives rise to two well known problems that cause performance degradations - Exposed-node (EN) and Hidden-node (HN) problems [4]. Both EN and HN are related to imperfect operation of the carrier-sensing mechanism in 802.11. Ideal carrier sensing must satisfy two criteria: (1) First, it should prevent simultaneous transmissions by interfering links. Otherwise, one or more of the transmissions will fail, resulting in retransmissions and bandwidth wastage. (2) Second, to exploit spectrum spatial reuse, it should allow simultaneous transmissions by non-interfering links. Prohibiting such simultaneous transmissions lowers the network throughput unnecessarily. Basically, HN arises when carrier sensing fails to satisfy (1); while EN arise when it fails to satisfy (2). Standard 802.11 networks often fail to satisfy the two criteria, hence their inherent HN and EN.

This paper is a first attempt for a comprehensive and rigorous study of EN and HN. There are three main contributions, as described below:

(1). Providing Rigorous Definitions for EN and HN
Previous studies of EN and HN have been based on incidental examples and specific topological layout of nodes. Rigorous definitions of EN and HN were lacking. Without such rigorous definitions, it would be difficult to devise comprehensive solutions to EN and HN and prove their validity under general settings. In addition, it would not be possible to measure the degrees of EN and HN in a network. This paper proposes formal definitions of EN and HN based on a graph model.

(2). Relating EN to Non-scalable Network Throughput and HN to Unfairness Throughput Distributions
It turns out that ultimately, it is EN that causes the sum of one-hop throughputs in 802.11 networks to be non-scalable. That is, the throughput problem in 802.11 WLAN cannot be solved by deploying more Access Points (APs): the overall network throughput quickly reaches a saturation point as more APs are added, and that EN is the underlying cause. HN does not cause non-scalability, but instead triggers other performance problems such as unfair throughput distributions and throughput instability [5].

(3). Investigating Solutions for Elimination of EN and HN respectively
This paper proposes and investigates solutions to eliminate EN and HN respectively. In particular, we show that it is possible to devise an 802.11-like CSMA/CA scheme to eliminate EN so that network throughput can scale according to the theoretical bounds established in [1] and [2]. In addition, it is possible to devise solutions to eliminate HN. However, there is generally a tradeoff between EN and HN, and entirely eliminating both of them together appears to be difficult.

Related Work

Previous investigations mostly addressed EN and HN separately. Among the studies of HN, many [6], [7], [8], [9] focus on solving HN-induced performance problems

- *Li Bin Jiang is with Dept. Electrical Engineering & Computer Science, University of California at Berkeley (e-mail: lj5@berkeley.edu).*
- *Soung Chang Liew is with Dept. Information Engineering, Chinese University of Hong Kong (e-mail: soung@ie.cuhk.edu.hk).*

Manuscript received 03 Sep, 2005; revised 11 Dec, 2006; accepted 05 Apr, 2007.

rather than elimination of HN itself. Compared with HN, there is relatively less work on EN. Specifically, the relationship between EN and scalability has not been formally established.

References [10] and [11] developed power-controlled MAC protocols to improve spatial reuse. Potentially, they can solve the non-scalability problem – neither explicitly examines the scalability issue, however. The scheme in [10] requires a node to continuously transmit busy tone signals on another channel in the form of short pulses while it is receiving a packet, and the scheme in [11] requires a separate control channel to exchange power information. These are substantial deviations from the current 802.11 standard and may not be simple to implement because of the complex transceiver designs that will be required. Reference [12] tried to remove EN and HN using "Dual Busy Tone Multiple Access". In that scheme, a node needs to transmit two narrow-bandwidth busy tones to notify its neighbors while receiving a signal. As with [10], the required frequency isolation in transceiver design is non-trivial. Unlike our work here, none of the previous work provides a systematic and comprehensive study of EN and HN. In addition, our solutions to EN and HN do not assume complex deviations (e.g., using control channel or busy tone channel, and/or requiring simultaneous receiving and transmitting on different channels) from the current 802.11-standard-based products in the transceiver design.

The rest of the paper is organized as follows. Section 2 explains HN and EN by example. Section 3 models the various constraints against simultaneous transmissions in terms of a set of mathematical inequalities. Section 4 formulates a graph model based on the constraints, and in so doing provides formal definitions for EN and HN. Section 5 presents a scheme called Selective Disregard of NAVs (SDN) to remove EN. Section 6 demonstrates the scalability of SDN and the non-scalability of original 802.11. Section 7 presents a scheme called Hidden-node Free Design (HFD) to remove HN. Section 8 shows that HFD removes known HN-related performance problems. Section 9 combines SDN and HFD in a complementary design to achieve the advantages of both. Finally, Section 10 concludes this paper.

2 Hidden-node and Exposed-node Problems

As mentioned in the introduction, ideal carrier sensing must satisfy two criteria: (1) It should prevent simultaneous transmissions by two interfering links. (2) It should allow simultaneous transmissions by non-interfering links. However, multi-access protocols, including 802.11, often fail to satisfy the above two criteria. HN occurs when the protocol fails to satisfy (1); while EN occurs when it fails to satisfy (2).

Now we give a few examples of HN and EN. Fig. 1 (a) depicts an 802.11 WLAN with an access point (AP) and two clients at distance d_{max} from the AP. When STA1 transmits to the AP, STA2 should not transmit to the AP, and vice versa. So, to avoid HN, the physical carrier-sensing range, *PCSRange*, should be larger than $2d_{max}$ so that the transmission of STA1 can be sensed by STA2, and vice versa.

The situation is different when there are multiple WLANs, as illustrated in Fig. 1 (b), in which *IR* is the "interference range" [17] for transmission over d_{max}. The ACK of AP1 to STA1 can destroy the DATA from AP2 to STA3, since AP1 is within *IR* of STA3. To avoid such collisions, we need $PCSRange \geq 2d_{max} + IR$. Generally, *IR* can be ex-

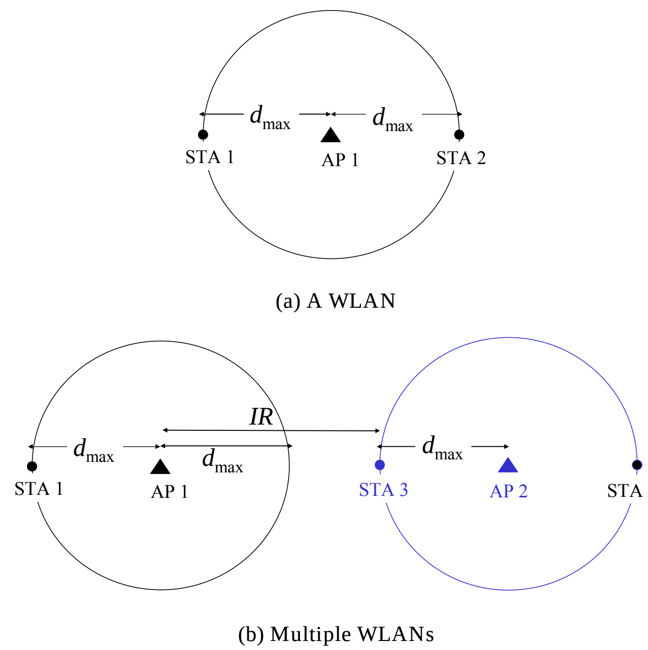


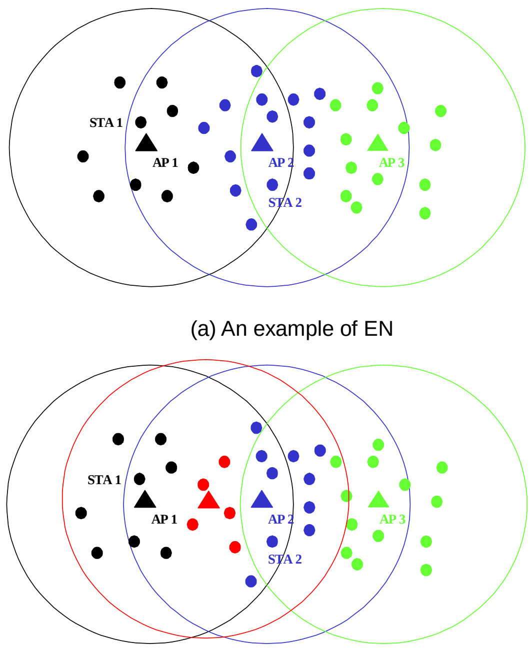
Fig. 1. Illustrations of hidden-node problem (HN).

pressed as $(1+\Delta)d$ for a link with distance *d*, where Δ is a positive distance margin related to the Signal-to-Interference Ratio (SIR) required for interference-free reception [17]. So, we need $PCSRange \geq (3+\Delta)d_{max}$. It can be proved this is one of the conditions that can ensure any network to be HN-free in general.

Fig. 2 (a) is an example of EN. Due to carrier sensing, links (AP1, STA1) and (AP2, STA2) cannot transmit together, even though there is no mutual interference, assuming $d(\text{AP1, AP2}) > (1+\Delta)\, d(\text{AP1, STA1})$ and $(1+\Delta)\, d(\text{AP2, STA2})$, where $d(i, j)$ is the distance between nodes *i* and *j*. However, $PCSRange > d(\text{AP1, AP2})$ prevents the simultaneous transmissions. In general, the larger the *PCSRange*, the worse is the EN problem.

As a result of EN, deploying more APs may not increase overall network throughput. Consider the example in Fig. 2 (b), where one AP has been added. The new AP does not help because all the client nodes "re-associated to" the new WLAN are still within the carrier-sensing range of the existing WLANs. Of course, one can use different frequency channels to isolate neighboring WLANs. However, frequency channels do run out when we add more and more APs (e.g., in 802.11b/g there are only three orthogonal frequency channels): eventually the overall network throughput will still become non-scalable as client stations increase.

3 Physical Interference Constraints and Protocol Constraints



(a) An example of EN

(b) Throughput is non-scalable due to EN

Fig. 2. Illustration of exposed-node problem (EN).

This section clarifies the relationships between link interferences and carrier sensing. Ultimately, it is the interplay between interferences and carrier sensing that is causing EN and HN. Throughout this paper, we assume all nodes use the same transmission power P_t. We define power-propagation function $P(P_t\ ;\ d)$ as the received power at distance d from the transmitter. In most cases, $P(P_t\,; d)$ will be simply expressed as $P(d)$. Links are modeled directionally so that the interconnections between two nodes are modeled as two directional links.

3.1 Protocol-independent Physical Interference Constraints

To model SIR-related physical interference, let T_i and R_i be the transmitter and receiver of link i. For brevity, we will also use T_i and R_i to denote their positions. Consider two links, 1 and 2. We assume that link 2 can interfere with link 1 if

$$P(|T_2 - R_1|) > P(|T_1 - R_1|)/K \qquad (1)$$

and link 1 can interfere with link 2 if

$$P(|T_1 - R_2|) > P(|T_2 - R_2|)/K \qquad (2)$$

where $K > 1$ is the minimum SIR required for proper detection. A typical value for K is 10. Simultaneous transmissions on links 1 and 2 will result in collisions if either (1) or (2) is true. (**NOTE**: In this paper, by "*simultaneous*", we refer to the situation when the any parts of the transmissions by different nodes overlap in time. Their transmissions may actually be initiated at different instances so that the start times of the transmissions are different.)

Note that (1) and (2) apply regardless of the multi-access protocol used. For a specific protocol, there may be additional interferences.

3.2 Protocol-specific Physical Interference Constraints

In 802.11, each atomic data transfer on a link i consists of a DATA frame in the forward direction followed by an ACK frame on the reverse link i' in the other direction. Not only is DATA-DATA collisions possible, the ACK on one link may collide with the DATA or ACK on other links. So, simultaneous communications on two links i and j are guaranteed to be interference-free only if the link pairs i and j, i and j', i' and j, and i' and j' are interference-free. Thus, collisions could also occur if any one of the following is true:

$$P(|R_2 - R_1|) > P(|T_1 - R_1|)/K \qquad (3)$$
$$P(|T_1 - T_2|) > P(|R_2 - T_2|)/K \qquad (4)$$
$$P(|T_2 - T_1|) > P(|R_1 - T_1|)/K \qquad (5)$$
$$P(|R_1 - R_2|) > P(|T_2 - R_2|)/K \qquad (6)$$
$$P(|R_2 - T_1|) > P(|R_1 - T_1|)/K \qquad (7)$$
$$P(|R_1 - T_2|) > P(|R_2 - T_2|)/K \qquad (8)$$

Inequalities (3) – (8) will be referred to as "protocol-specific physical-interference constraints", because these interferences would not be there if not for the protocol-specific ARQ mechanism.

We note that (1) – (8) can be easily transformed to distance relationships of the type $|C - B| < (1+\Delta)|A - B|$, where ($A$, B) is the link being interfered by C [1][13], if the power-propagation function $P(d)$ is a well defined non-decreasing function of the distance d.

3.3 Protocol Collision-Prevention Constraints in 802.11

There are two subtypes of "Protocol Collision-Prevention Constraints" in 802.11, as follows.

Transmitter-Side Carrier–Sensing Constraints

Carrier sensing can be used to avoid simultaneous transmissions that collide. Ideally, it should make use of (1) – (8) to decide which simultaneous transmissions are allowed. However, 802.11 imposes a set of carrier-sensing constraints that are distinct from (1) – (8).

There are two carrier-sensing mechanisms in 802.11: virtual carrier-sensing (VCS) and physical carrier-sensing (PCS) [3]. With VCS, a node keeps silent in intervals specified by the NAV ("Network Allocation Vector") information in RTS/CTS frames received from other nodes. With PCS, a node keeps silent when it senses enough strong power in the medium, or in intervals specified by the length field in the PHY headers of packets received from other nodes. For generality, we assume *RTS/CTS Access Mode* in the following discussion, where both VCS and PCS are operational. For *Basic Access Mode*, inequalities (9), (10), (12) and (13) below could be eliminated.

We assume RTS/CTS can be decoded if the received power is larger than a threshold P_v, or equivalently, the transmission distance of RTS/CTS is less than the virtual carrier sensing range, *VCSRange*. Consider links 1 and 2. Suppose that link 1's transmission is already in progress when link 2 has a packet to transmit. Link 2 cannot transmit if

$$|T_2 - T_1| < VCSRange\ \ OR\ \ |T_2 - R_1| < VCSRange \qquad (9)$$

If (9) is satisfied, T_2 will have received the RTS of T_1 or the CTS of R_1, and the NAV contained in the RTS or CTS will prevent T_2 from transmitting. Likewise, suppose link 2's transmission starts first. Then link 1 cannot transmit if

$$|T_2 - T_1| < VCSRange \ \ OR \ \ |R_2 - T_1| < VCSRange \qquad (10)$$

If T_1 and T_2 can physically carrier-sense frames transmitted by each other, simultaneous transmissions will also be prevented. This is the case if the received power is larger than a threshold P_P, which can be mapped to a physical carrier-sensing range, *PCSRange*:

$$|T_2 - T_1| < PCSRange \qquad (11)$$

Although T_1 will not transmit if it senses the ACK from R_2 (and similarly for T_2 and R_1), we ignore these constraints because the "air-time" of ACK is usually much smaller than DATA.

Inequalities (9)-(11) will be referred to as "Transmitter-Side Carrier-Sensing Constraints". They prevent a transmitter from transmitting when it already senses another transmission. If the transmission would indeed result in a collision had it gone ahead, then the transmitter has made the right decision. However, there are situations when a transmitter-side carrier-sensing constraint is satisfied while none of the physical-interference constraints is satisfied. In that case, the transmitter would have made a wrong decision. This is an EN situation. On the other hand, even if simultaneous transmissions are allowed by (9) – (11), a collision may result if any of the inequalities (1) – (8) is satisfied. This is an HN situation.

Receiver-Side Carrier Sensing Constraints

Another HN situation may arise due to receiver-side carrier sensing. Specifically, the later one of two overlapping transmissions may fail if one of the following is true:

$$|R_2 - T_1| < VCSRange \ \ OR \ \ |R_2 - R_1| < VCSRange \qquad (12)$$

$$|R_1 - T_2| < VCSRange \ \ OR \ \ |R_1 - R_2| < VCSRange \qquad (13)$$

$$|R_2 - T_1| < PCSRange \qquad (14)$$

$$|R_1 - T_2| < PCSRange \qquad (15)$$

Inequalities (12) – (15) will be referred to as "Receiver-Side Carrier Sensing Constraints". When none of the "Transmitter-Side Carrier-Sensing Constraints" ((9)-(11)) is true, T_1 and T_2 are allowed to transmit simultaneously. However, if any of (12) to (15) is satisfied, the receiver will not reply a CTS/ACK to the transmitter, causing the transmitter to interpret that as a collision, which then triggers backoff and retransmissions.

Consider (12) (similar argument applies to (13)). If link 1 starts its transmission first, then R_2 has heard the RTS or CTS from T_1 or R_1, if (12) holds. When T_2 attempts to initiate a transmission later by sending an RTS to R_2, R_2 will not reply with a CTS since its NAV has been set. Inequalities (14) and (15) are due to the default operation mode of the receiver design in most 802.11 products today. This operation mode is also assumed in the NS2 simulator [14]. Consider (14) (similar argument applies to (15)). If T_1 starts its transmission first followed by T_2, then R_2 will have locked on to the reception of the signal from T_1 and it will not attempt to receive the overlapping signal from T_2 that arrives later, even if the signal from T_2 is much stronger. A rationale for this design could be that if R_2 were to receive the signal from T_2 and then return an ACK, this ACK might interfere with the transmission on link 1.

Note that (12) – (15) are conditions that lead to the classical HN situation [5][9]. At the same time, (12) – (15) are also conditions that *may* lead to an EN situation: this will be the case when none of (1) – (8) is satisfied. In that case, there is actually no interference. In the above example, R_2 can safely reply to T_2 without fearing that it will cause interference on link 1. However, because it is exposed to the signal from link 1, it will refrain from doing so. The more precise definitions for EN and HN will be given in the next section.

4 FORMAL DEFINITIONS OF EN AND HN

This section provides formal definitions for EN and HN based on a graph model derived from inequalities (1) – (15) and demonstrates the tradeoff between EN and HN. Let us assume the "two-ray ground" propagation model [15] [16]. The received power function is

$$P(d) \propto P_t / d^{\alpha}$$

where P_t is the transmission power, d is the distance, and α is the path-loss exponent. If K=10, and $\alpha = 4$, then from [17] the interference range (IR) due to the physical constraints on link (T_A, R_A) is $IR = K^{1/\alpha} d_A = 1.78 d_A$ where $d_A = |T_A - R_A|$. That is, there should not be another active transmitter within the IR of R_A when T_A is transmitting to R_A. In particular, IR depends on the distance d_A. The *VCSRange* and *PCSRange* for the protocol constraints (9) – (15), on the other hand, are independent of d_A. Thus, 802.11 can not approximate the physical constraints well.

4.1 Link-Interference Graph from Physical Interference Constraints

A link-interference graph (i-graph) can be used to capture the physical interference constraints graphically. In an i-graph, a vertex represents a wireless link. This is an i-edge between vertices 1 and 2, if any of the inequalities (1) - (8) is satisfied. An i-edge means link 1 physically interferes with link 2, or vice versa. We define i-edges to be *non-directional* here.

We next define an s-graph. There is a *directional* s-edge from vertex 1 to vertex 2, and a directional s-edge from vertex 2 to vertex 1 when there is an i-edge between the two vertexes. Interference in either direction implies s-edges in both directions. The interpretation of s-edge is as follows. **An s-edge from vertex 1 to vertex 2 means that in order to prevent future collisions, link 1 must be capable of forewarning link 2 not to transmit when link 1 initiates a transmission.** Whether the interference is from link 1 to link 2, or from link 2 to link 1, link 1 must be able to do so. The vertices and the s-edges constitute the s-graph.

Fig. 3 (a) and (b) show an example of mapping a network topology to an s-graph. In Fig. 3 (a), assume link 2 can interfere with link 1, but not vice versa (because of the shorter length of link 2). There is an i-edge between their vertexes according to our definition. To prevent link 1's from being interfered by link 2, when link 1 transmits, it should forewarn link 2 not to transmit. Therefore, there is an s-edge

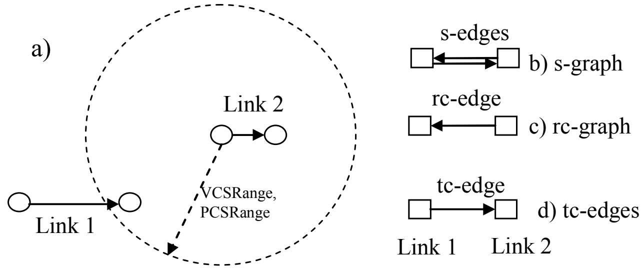


Fig. 3. Mapping of a network topology a) to b) s-graph and c) rc-graph and d) tc-graph. (For convenience, assume VCSRange = PCSRange in this Figure)

from vertex 1 to vertex 2. Likewise, when link 2 transmits, it should forewarn link 1 not to transmit, else the signal on link 1 will be corrupted anyway. Therefore, there is an s-edge from vertex 2 to vertex 1. Note that s-edges represent what SHOULD HAVE happened. The actual carrier-sensing operation may fail to achieve this, however.

4.2 Transmitter-Side Carrier–Sensing Graph (tc-graph)

The tc-graph is to model the Transmitter-Side Carrier–Sensing. For 802.11, there is a *directional* tc-edge from vertex 1 to vertex 2 if the (9) is true; and there is a directional tc-edge from vertex 2 to 1 if (10) is true. There are two tc-edges (1→2 and 2→1) if (11) is satisfied. A tc-edge 1→2 means that link 1 can and will forewarn link 2 not to transmit.

Fig. 3 (a) and (d) show an example of mapping a network topology to a tc-graph. In this example, the transmitters of links 1 and 2 cannot hear each other. However, the transmitter of link 2 can hear the CTS sent by the receiver of link 1. So, there is a tc-edge from link 1 to link 2. However, the transmitter of link 1 can neither hear the transmitter nor receiver of link 2. Hence, there is no tc-edge from link 2 to link 1.

4.3 Receiver-Side Carrier–Sensing Graph (rc-graph)

The rc-graph is to model the Receiver-Side Carrier-Sensing Constraints. For 802.11, there is an rc-edge 1→2 if (12) or (14) is true; and an rc-edge 2→1 if (13) or (15) is true. As with the tc-graph, the directional edges in rc-graph may not be symmetric. An example is shown in Fig. 3 (a) and (c).

The sets of edges in s-graph, tc-graph and rc-graph are three different sets, as shown in the Venn Diagram in Fig. 4.

4.4 Formal definitions of EN and HN

The phenomena of exposed and hidden nodes are due to relationship between links rather than that between nodes. So, exposed- and hidden-node problems could be more accurately defined as *expose*- and *hidden-link* problems. However, in view of the fact that the terms “exposed and hidden nodes” have already been widely used by the research community, we will continue to adopt the less accurate terms.

Definition of EN. *EN exists from link i to link j if and only if there is a tc-edge or rc-edge, but no s-edge, from vertex i to vertex j. Link j is said to be exposed to link i in this case.*

The interpretation is as follows. Since there is no s-edge from vertex *i* to vertex *j*, there is no need for link *i* to forewarn link *j* when link *i* transmits because they do not mutually interfere. However, the carrier-sensing operation will either prevent link *j* from transmitting (in the case of a tc-edge), or prevent the success of the transmission (in the case of no tc-edge, but an rc-edge). In Fig. 5 (a), the green part shows the set of edges which cause EN.

Definition of HN. *HN exists from link i to link j if and only if there is an s-edge or rc-edge, but no tc-edge, from vertex i to vertex j. Link i is said to be hidden from link j in this case.*

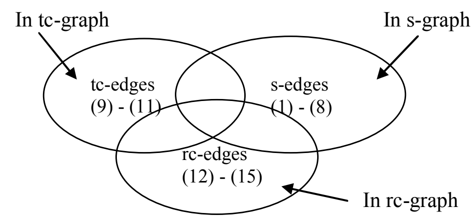
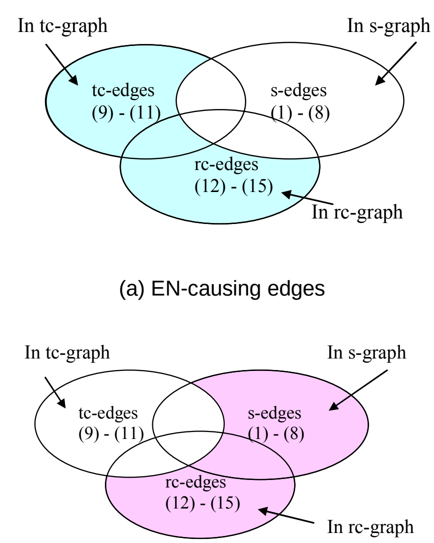


Fig. 4 Relationships between edges in s-graph, tc-graph and rc-graph, and the inequalities associated with the edges. For any vertex pair *i* and *j*, (*i*, *j*) could be an s-edge, tc-edge, rc-edge or none of them. Note that some (*i*, *j*) may belong to s-graph, tc-graph, and rc-graph simultaneously

The interpretation is as follows. Since there is no tc-edge from vertex *i* to vertex *j*, the transmitter of link *j* will not be able to carrier-sense the transmission on link *i*. However, there is either physical interference from *i* to *j*



(b HN-causing edges

Fig. 5. The edges causing EN and HN.

(in the case of an s-edge), or the receiver of *j* will ignore the transmitter of *j* (in the case of an rc-edge) when the receiver already senses the transmission on link *i*. In either case, the transmitter of link *j* will interpret that as a collision. The carrier-sensing operation has failed in this case because of HN. In Fig. 5 (b), the red area shows the set of edges which lead to HN.

According to the above definitions, in the example topology of Fig. 3 (a), there is no EN, but there is HN: link 2 is hidden from link 1.

EN and HN are caused by *discrepancies* among s-graph, rc-graph and tc-graph. Denote the set of tc-edges in the tc-graph by *TC*, denote the set of s-edges in the s-graph by *S*, and denote the set of rc-edges in the rc-graph by *RC*. We can compute the following measures of interest:

- # of HN-causing edges: $N_{HN} = |(S \cup RC) \cap \overline{TC}|$.
- # of EN-causing edges: $N_{EN} = |(TC \cup RC) \cap \overline{S}|$.
- MISS RATIO= $N_{HN} \,/\, |S \cup RC|$ (ideally, *TC* needs to fully cover $|S \cup RC|$ to remove HN).
- FALSE-ALARM RATIO= $N_{EN} \,/\, |S|$ (to remove EN completely, $TC \cup RC$ should not go beyond *S*).

Note that while EN and HN have been defined as *local* relationships between two links, miss ratio and false-alarm ratio are *global* measures of the severities of HN and EN in the overall network.

We now study the tradeoff between EN and HN assuming the use of 802.11b, which supports four data transmission rates, 1, 2, 5.5 and 11Mbps. Table 1 shows the maximum transmission ranges (*TxRange*), adopted from [18].

We generated three random network topologies with 100 nodes in a circular plane with 2km radius and compute their respective s-graph, rc-graph and tc-graph for each setting. A link is formed by a pair of nodes: one node acts as the transmitter and the other as the receiver. And for convenience, each node only belongs to one link, where actual data transmission takes place. So there are altogether 50 links in each topology. Four cases corresponding to different *TxRange*, *VCSRange*, *PCSRange* settings are considered and they are listed in Table 2. The miss ratio and false-alarm ratio are shown in Table 3.

Both HN and EN exist in all the four cases. Generally, changing the *PCSRange*, *VCSRange* or *TxRange* in 802.11 cannot eliminate HN and EN entirely. There is also a tradeoff between HN and EN. The severity of HN and EN depends on various factors, such as the range settings, the node density, specific topologies and traffic patterns.

TABLE 1
TRANSMISSION RANGES FOR VARIOUS DATA RATES

DATA RATE (MBPS)	1	2	5.5	11
TXRANGE (M)	550	437	292	232

TABLE 2
4 EXAMPLE CASES OF TXRANGE, VCSRANGE AND PCSRANGE SETTINGS

	TXRANGE (M)	*VCSRANGE* (M)	*PCSRANGE* (M)
CASE 1	437	437	437
CASE 2	232	437	437
CASE 3	232	437	550
CASE 4	232	232	550

TABLE 3
HN-CAUSING EDGES AND EN-CAUSING EDGES (AVERAGED OVER 3 RANDOM TOPOLOGIES)

	# of HN-causing edges	MISS RATIO (%)	# of EN-causing edges	FALSE-ALARM RATIO (%)
Case 1	172	61.9	5.33	1.96
Case 2	23.7	19.8	55.3	69.4
Case 3	28.0	16.6	104	116
Case 4	26.3	16.8	78.3	88.1

5 SELECTIVE DISREGARD OF NAVS (SDN)

Sections 5 and 7 present a scheme to remove EN and a scheme to remove HN, and Sections 6 and 8 examine their performance, respectively. To remove EN, we need to find a way to carrier-sense without limiting spatial reuse. That is, we need to eliminate unnecessary tc-edges represented by inequalities (9)-(11) and rc-edges represented by (12)-(15). To remove HN, on the other hand, we should ensure interfering links can carrier-sense each other. This section presents a scheme called Selective Disregard of NAVs (SDN) to remove EN. *By itself, SDN does not try to remove HN*. There are three parts to SDN, described as follows.

SDN. I - Turning off PCS and Using RS

To achieve scalable performance, physical carrier sensing needs to be deactivated, so that a node's transmission decision depends only on the NAVs, as described in SDN.II below. This means that a node can initiate a transmission as long as its NAVs allow it, whether or not the medium is sensed as busy physically. Deactivating physical carrier sensing removes (11).

In some commercial 802.11 chips, there is a receiver "**Restart mode**" (RS) that is available **at the physical layer** (i.e., below the 802.11 MAC layer), in which a receiver will switch to receive a stronger signal in the midst of receiving a weaker signal if the power difference is sufficiently large (usually, the ratio is the minimum SIR ratio required for detection, *K*). This feature can be used to lift the limitations imposed by inequalities (14) and (15). Then, simultaneous transmissions of two links will succeed as long as there is no i-edge between them.

By removing (11), (14), and (15), the associated tc- and rc- edges have also been removed. The deactivation of physical carrier sensing may potentially cause HN. HN can be effectively reduced by incorporating a “Hidden-Node Free design”, which will be discussed later.

SDN.II - Selective Disregard of NAV (SDN)

SDN.II is to ensure that VCS **allows** simultaneous transmissions that satisfy none of (1) – (8). The VCS of 802.11 is modified so that the NAV in RTS/CTS packets is only selectively considered. Each node not only monitors whether there are other ongoing transmissions, but also who is transmitting to whom (by reading the transmitter and receiver addresses in the RTS/CTS it has received). A node may transmit its own frame provided there is no i-edge between its link and the currently transmitting links. In the 802.11 standard, RTS contains both addresses, but CTS only contains the receiver address, so a new field containing the transmitter address should be added to CTS. Together with SDN.I, SDN.II removes the effects of edges in $(TC \cup RC) \cap \bar{S}$. The details of the SDN.II algorithm are as follows:

(i) Each node a keeps an NAV set, NAVa = {NAV(a, k)}, consisting of the NAVs node a hears in its neighborhood, where k is the label for links. NAVs heard are contained in the RTS and CTS sent by other nodes.

(ii) Suppose node a has a packet to transmit to node b. Node a will send an RTS to b if and only if for all k with NAV(a, k) > 0, there is no s-edge between vertex (a, b) and vertex k. Note that unlike in the standard 802.11, NAVs from non-interfering links are ignored.

(iii) Node b also keeps an NAV set, NAVb={NAV(b, m)} that b hears in its neighborhood, where m is the label for links. After receiving node a's RTS, it will reply with a CTS if and only if there is no s-edges between vertex (a, b) and vertex m.

An assumption in (ii) is that a node knows the s-edges in its neighborhood: the power exchange (PE) algorithm in SDN.III is used to construct the s-edges in a distributed manner. With (ii), a transmission will proceed only if (a) doing so will not cause interference to the current transmissions; and (b) the transmission will not be interfered by the current transmissions. Note that the selective nature of SDN.I and II gets rid of rc-edges and tc-edges where there is no s-edge.

Note that SDN can operate without removing HN entirely, as the original 802.11 networks can operate while there is HN. However, even if we do not want to remove HN entirely, to avoid excessive collisions, a reasonable *VCSRange* should be larger than the maximum Interference Range, IR_{max}, which satisfies $P(IR_{\max}) = P(d_{\max})/K$.

SDN.III - Constructing s-graph using PE Algorithm

In SDN.II, each node needs to know the s-edges of its links. To construct the s-edges, each node a periodically broadcasts special power-exchange (PE) packets and receives PE packets from nearby nodes. The periodical broadcasting is for robustness and more accurate estimation of power. The transmission power for PE packets is the same as that for regular packets, and we assume that all nodes use the same transmit power and receiver sensitivity.

(i) Node a measures the powers of PE packets transmitted by other nodes that it can hear, and keeps the power information in a "power set", P_a = {[c, $P(c, a)$]} where c is the node label of the sender of the PE packets.

(ii) Periodically, node a broadcasts one PE packet, which contain (1) A list of active links (a, b) or (b, a) (b is any other node forming a link with a); and (2) P_a. (We do not require the periodic broadcasts of different nodes to be synchronous).

(iii) Node a identifies its associated s-edges based on the PE packets that it receives (The detailed procedure is described in the Appendix I.)

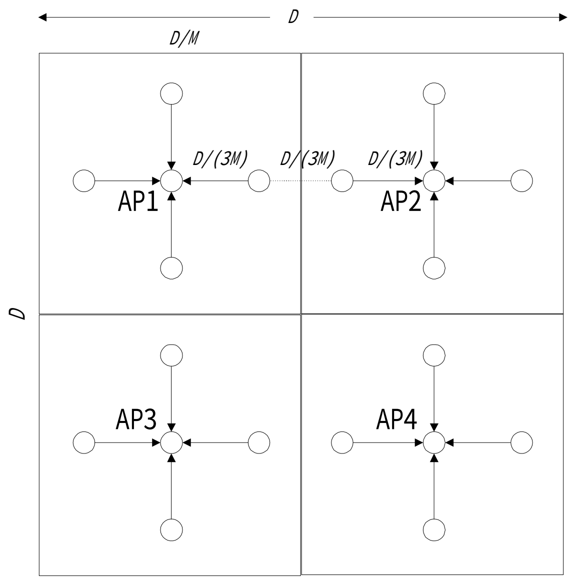
Condition for Correct Operation of PE

The following condition is sufficient to ensure a node can construct its associated s-edges:

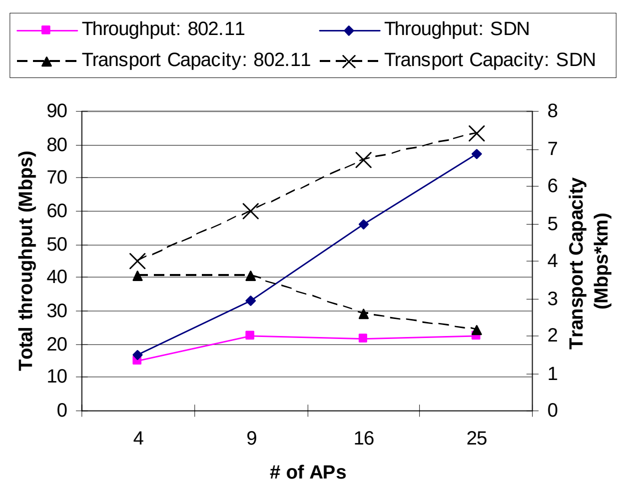
$$P(PERange - d_{\max}) < P(d_{\max})/K \tag{16}$$

where d_{max} is the maximum link distance in the network, *PERange* is the transmission range of the *PE packets*, and P(.) is the received power as a decreasing function of distance. The proof is given in the Appendix I.

A point to note about the PE algorithm is as follows. PE packets are not the same as RTS/CTS packets, and are not used for carrier-sensing purposes. They are special packets used for distributed construction of s-edges. They are transmitted periodically for robustness and accurate estimation of power, especially when the network topology or conditions have changed. In the indoor environ-

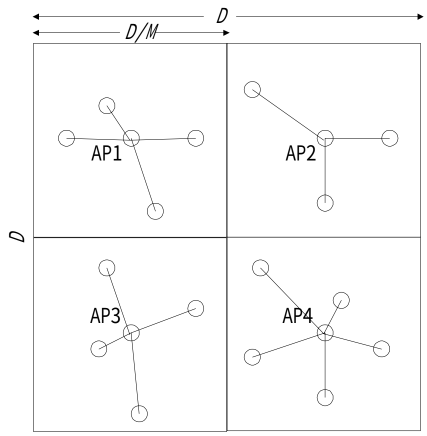


(a) A multi-cell Infrastructure Network with regular client-node placement (when M=2)

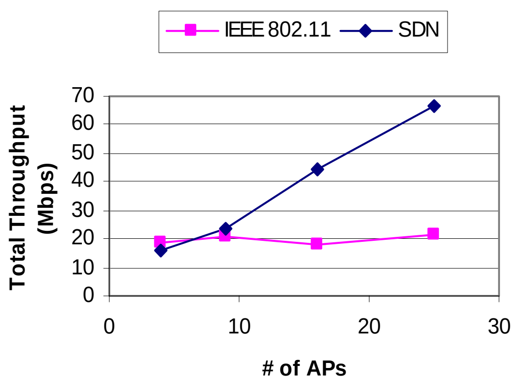


(b Total throughputs (Mb/s) and Transport Capacity (Mb/s*km)

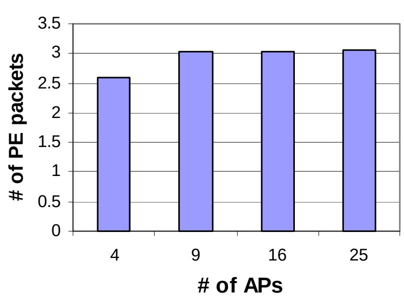
Fig. 6 Comparison of 802.11 and SDN in a regular topology.



(a) A multi-cell Infrastructure Network with random client-node placement (here, M=2)



(b) Total throughputs (Mb/s)



(c) Average number of PE packets transmitted per node before the initial convergence

Fig. 7 Comparison of 802.11 and SDN in random topologies.

ments where node movement and channel variation is slow, the overhead of PE algorithm is low since the broadcast period can be long. Even if more frequent broadcast is needed, PE algorithm is still worthwhile considering its benefit—achieving O(n)-scalable (instead of O(1)) throughput by addressing the EN problem. The benefit increases unboundedly when n increases. This will be elaborated in the next section and Appendix II.

6 EN AND ITS IMPACT ON SCALABILITY

This section investigates the impact of EN on scalability. We focus on infrastructure networks rather than *ad hoc* networks here. Analysis of *ad hoc* networks can adopt a similar approach with similar qualitative results. We show by simulations and analytical arguments that, in terms of the sum of one-hop throughputs, the original 802.11 is non-scalable. SDN, on the other hand, is scalable and can achieve $O(n)$ throughput where n is the number of nodes per unit area. This matches the theoretical results under "perfect scheduling" in [2] as far as the order of throughput is concerned.

Now we validate the performance of SDN by NS-2 simulations. Consider the grid topology in Fig. 6 (a). In a square area of $D*D$, there are M^2 equal-sized cells. In each cell, 4 client stations are associated with an AP at the center, forming an Infrastructure Network. There is a saturated UDP flow from each client to its AP. The length of a cell's edge is D/M, and the distance between each client and its AP is $D/(3M)$. D=1440m, $PCSRange$=550m, and $VCSRange$=437m. The power margin K=10. And "two-ray ground" propagation model is adopted. We progressively let M = 2, 3, 4 and 5. Since there are totally M^2 APs and $4M^2$ clients in the area, the node density increases with M.

The total throughputs ($\sum_i T_i$, where T_i is the throughput of link i) and transport capacities ($\sum_i T_i d_i$, where d_i is the distance between the transmitter and receiver of link i [1]) with SDN and original 802.11 are shown in Fig. 6 (b).

In Fig. 7 (a), the clients are *randomly* located with a uniform distribution. Each client establishes a link with the nearest AP, and (# of clients) / (# of APs) remains 4. The total throughputs with SDN and original 802.11, averaged over three random topologies, are shown in Fig. 7 (b).

From Fig. 6 (b) and Fig. 7 (b), we see that because of EN, the total throughput of 802.11 reaches a limit when the # of APs increases, revealing its non-scalability. When more WLANs and user nodes are added so that the network becomes more densely populated, the aggregate throughput per WLAN decreases. SDN effectively solves the scalability problem - its total throughput is nearly proportional to the number of APs, and the throughput per WLAN does not decrease. Also, unlike SDN, the "transport capacity" of 802.11 in Fig. 6 (b) decreases with the increase of the # of APs, since the throughput does not increase while the link distances decrease.

About the performance of PE algorithm, Fig. 7 (c) shows the average number of PE packets transmitted (per node) before all nodes initially constructed their s-edges. Ideally, the time needed for this initial convergence should be within two periods, since each node needs two-hop information (see Appendix I). But due to packet collisions, the average # of PE packet per node is higher than 2. After the initial convergence, each node continues broadcasting PE packets periodically for robustness and to adapt to channel variations. In indoor environments where node movements and channel variations are slow, the broadcast period can be made, say, 1 sec, which leads to little overhead. Even when more frequent broadcasts are needed, the overhead cannot offset the throughput gain from O(1) of 802.11 to O(n) of SDN.

Appendix II gives the analytical arguments for the above observations. That is, SDN is scalable—achieving O(n) total throughput; while 802.11 is unscalable—achieving O(1) total throughput.

7 Hidden-node Free Design (HFD)

Recall that HN exists between two links if 1) transmissions on the two links may interfere with each other to cause reception failure on one or both links; and 2) the *sender* of one of the links cannot sense the transmission on the other link, and/or vice versa. According to the formal definition of HN (Section 0), the set of edges which causes HN is $(S \cup RC) \cap \overline{TC}$. Therefore to remove HN, we should guarantee that

$$(S \cup RC) \cap \overline{TC} = (S \cap \overline{TC}) \cup (RC \cap \overline{TC}) = \varnothing \tag{17}$$

Equivalent to (17), the following two conditions should be guaranteed:

$$S \cap \overline{TC} = \varnothing \tag{18}$$

$$RC \cap \overline{TC} = \varnothing \tag{19}$$

We now present a set of *sufficient conditions* for removing HN, which we refer to as Hidden-node Free Design (HFD). Note that HFD does not necessitate SDN, and vice versa, although they can be combined to solve HN and EN in a complementary way. The treatments in Section 7 and Section 5 are parallel and independent. In Section 9 we will combine SDN and HFD.

7.1 HFD for IEEE 802.11 Basic Access Mode

We first consider HFD for the IEEE 802.11 Basic Access Mode. The basic idea is to

(i) Make RC empty (satisfy (19)), and
(ii) Make TC fully cover S (satisfy (18)).

HFD for Basic Access Mode [20] consists of (a) the Restart Mode (to achieve (i)); and (b) a constraint on the power budget of links (to achieve (ii)).

(a) Receiver Restart Mode (RS): As has been mentioned in Section 5, this is a receiver mode that can be enabled in some commercial chips. Like normal receivers without RS, a receiver with RS that has sensed a transmission in progress should not initiate a new DATA transmission of its own. Also, if it receives a new signal with less than $1/C_t$ times the power of the previous signal, the receiver will properly decode the previous signal and ignore the later signal. However, with RS, if the power of the new signal is more than C_{rt} times the power of the previous signal, the receiver will then *switch* to receive the stronger new signal. For simplicity, we assume $C_{rt} = C_t = K$.[1] If the new signal is a DATA targeted for the receiver, the receiver will then reply with an *ACK*. Note that receivers without RS *will not attempt* to receive the new signal even if it is much stronger than the previous signal. That behavior is called "receiver capture".

(b) Link Power-Budget Requirement: In addition to (a), the following inequality needs to be fulfilled to eliminate HN:

$$P(d_{\max}) \geq C_t P(PCS - 2d_{\max}) \tag{20}$$

where P(.) is the received power as a function of distance, d_{max} is the maximum link length in the network, PCS is the physical carrier-sensing range, and C_t is the detection threshold. An implicit assumption in the above inequality is that P(.) is a decreasing function of distance.

If we define a "maximum Interference Range" IR_{max}, which satisfies

$$P(IR_{\max}) = P(d_{\max}) / C_t \tag{21}$$

then (20) is equivalent to

$$PCS \geq 2d_{\max} + IR_{\max} \tag{22}$$

This has been graphically explained in Section 2 (see Fig. 1 (b)). Note that this is a requirement imposed on the network design. PCS should be large enough relative to the maximum link length d_{max} if HN is to be removed.

We could plug in a suitable propagation model to the above requirements. The received power function is usually in the form of

$$P(d) \propto P_t / d^{\alpha} \tag{23}$$

where P_t is the transmission power, d is the distance and α is the path-loss exponent, which ranges from 2 to 6 according to different environments [16]. For example, assuming C_t=C_{rt}=10, $\alpha = 4$, and defining $\Delta = C_t^{1/\alpha} - 1$, the requirement then becomes

$$PCS \geq 2d_{\max} + (1+\Delta)d_{\max} \approx 3.78d_{\max} \tag{24}$$

In [19][20], we have given detailed proof that (a) and (b) ensure that the network is HN-free. We omit the proof here to conserve space.

Implementation Considerations

To ensure (24), we can either adjust PCS (PCS Range), or imposed a limit on the maximum link length d_{max} such that it is sufficiently smaller than the maximum transmission range, or both. For HFD for RTS/CTS in the next subsection, we also need to adjust VCS (VCS Range), so we will discuss the implementation issues together here. PCS, VCS and d_{max} are determined by transmission power P_t and receiving thresholds. The following discussion assumes all nodes use the same uniform P_t and receiving thresholds, hence they have the same PCS, VCS and d_{max}. We can write

$$P(P_t; d_{\max}) = P_{link} \tag{25}$$

$$P(P_t; PCS) = P_{PCS} \tag{26}$$

$$P(P_t; VCS) = P_{VCS} \tag{27}$$

where $P(P_t; x)$ is the received power at a distance x; P_{link} is the received power threshold required to establish a link – we do not set up a link if the received power falls below this threshold; P_{PCS} is the received power threshold for physical carrier sensing – if the received power is lower than P_{PCS}, PCS will not be operated even if the power is sufficient to decode the PHY header; and P_{VCS} is the received power threshold for virtual carrier sensing – no NAV will be set even if the RTS/CTS packets can be decoded if the received power falls below this threshold. In general, $PCS > VCS > d_{max}$ if excessive HN collisions are to be avoided, which means $P_{PCS} < P_{VCS} < P_{link.}$ .This translates to the following receiver carrier-sensing operation:

1. The receiver will only attempt to decode a signal if the received power $P_r > P_{PCS}$.

[1] In SDN (Section 5 and 6), we have also assumed $C_{rt} = C_t = K$.

2. If $P_r > P_{PCS}$, the receiver will first attempt to decode the PHY header. If the PHY header cannot be decoded (could be a non-802.11 source, or 802.11 collided packets), then *power carrier-sensing* kicks in – the receiver will continue to regard the medium as "busy" until the power level falls below P_{PCS}.
3. If the PHY header can be decoded but not the MAC payload, then wait until the end of the packet (can be deduced from the length field in the PHY header) plus EIFS [3] in accordance to the 802.11 specification before regarding the medium as "idle".
4. If the PHY header and MAC payload can be decoded, and if this is an RTS/CTS packet, set the NAV in accordance to the 802.11 specification only if $P_r > P_{VCS}$; do nothing otherwise.
5. If the PHY header and MAC payload can be decoded, and this is a regular DATA/ACK packet targeted for others, operate in accordance to the 802.11 specification.

PCS, *VCS* and d_{max} could be adjusted by tuning the power thresholds. For example, if we want PCS=3.78d_{max}, with the assumption of two-ray ground model, (25) and (26) become

$$A \cdot P_t \cdot d_{\max}^{-4} = P_{link} \tag{28}$$

$$A \cdot P_t \cdot PCS^{-4} = P_{PCS} \tag{29}$$

where A is a constant [16] . So, we have

$$P_{link} = 3.78^4 P_{PCS}, \text{ or } P_{PCS}(dB) \approx P_{link}(dB) - 23.10(dB) \tag{30}$$

But clearly, these thresholds cannot be infinitely small due to physical-layer feasibility. If the receiver sensitivities needed to decode DATA, PHY header, and RTS/CTS frame (which are transmitted at different bit rates and therefore have different SIR requirements) are P_l, P_P and P_V, then we must ensure $P_{link} \ge P_l, P_{PCS} \ge P_P, P_{VCS} \ge P_V$. If these inequalities cannot be satisfied then it will be necessary to increase P_t.

The above condition PCS=3.78d_{max} seems to be stringent. However, they are fundamental conditions we have identified in order to remove HN entirely. In practice, however, with specific design scenarios and limitations, one may choose to relax the requirement if certain amount of HN can be tolerated. Our conditions here provide a useful guideline. (In particular, **Fig. 13** will plot a series of tradeoff points a designer can choose from.)

7.2 HFD for IEEE 802.11 RTS/CTS Access Mode

For RTS/CTS Access Mode, *TC* contains the edges specified by (9) - (11), and *RC* contains the edges specified by (12) - (15). HFD for RTS/CTS mode consists of four parts:

(i) PCS for RTS/CTS is turned on, but PCS for DATA/ACK is turned off - this means a node will not refrain from transmission if it senses a DATA/ACK packet. For implementation, a new bit can be added to the PHY header to indicate whether the packet type is RTS/CTS or DATA/ACK. After the PHY header is decoded, a node decides whether to operate PCS. On the other hand, if the PHY header cannot be decoded due to interference, then power carrier-sensing will kick in as a conservative measure (see item 2 in the carrier-sensing operation above).

(ii) Receiver Re-start Mode, to remove (14) and (15) from *RC*.

(iii) When a node receives an RTS targeted for it, it will always reply with a CTS regardless of NAV. Note that this is similar to the behavior of ACK (no carrier-sensing is needed before sending an ACK). Its purpose is to remove (12) and (13). With (ii) and (iii), the set *RC* becomes empty.

(iv) If $C_{rt} = C_t$, two power budget requirements as described in the following:

Power budget requirement 1:

$$P(d_{\max}) \ge C_t P(VCS - d_{\max}), \tag{31}$$
$$\text{or equivalently, } VCS \ge d_{\max} + IR_{\max}$$

where IR_{max} satisfies (21).

Consider two links, 1 and 2, which mutually interfere because an inequality in (1) - (8) is true. Then, to prevent collision and HN, T_1 and T_2 must be able to hear the CTS **or** RTS transmitted on the other link. It's not difficult to verify that inequality (31) ensures this [19].

Power budget requirement 2:

$$P(d_{\max}) \ge C_t P(PCS_{RTS/CTS} - 2d_{\max}), \tag{32}$$
$$\text{or equivalently, } PCS_{RTS/CTS} \ge 2d_{\max} + IR_{\max}$$

That is, the *PCS* Range of RTS/CTS should cover the *sender* of an interfering link (similar to (20) in HFD for Basic Access Mode). Its purpose is to avoid collisions among RTS/CTS, which carries important NAV information. Note that unlike DATA/ACK, RTS/CTS packets can only be protected by the PCS mechanism, but not the VCS mechanism. The proof of (32) is similar to the proof of HFD for basic Access Mode.

The difference between inequalities (32) and (31) is as follows. Inequality (32) is to guarantee that two interfering links can warn off each other of potential RTS/CTS collisions through PCS *before* the RTS/CTS can be decoded, and it addresses HN collisions among RTS/CTS. Inequality (31), on the other hand, is to guarantee that any two interfering links can warn off each other through VCS *after* RTS/CTS are decoded, avoiding DATA/ACK collisions.

8 Performance Evaluation of HFD

We only present performance results of HFD for Basic Access Mode here. The results of HFD for RTS/CTS Mode are similar qualitatively and can be found in [19].

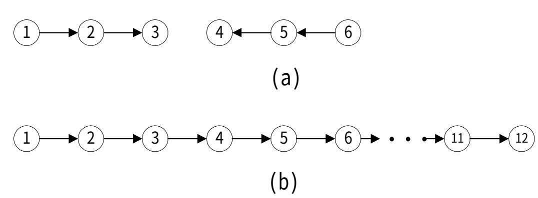
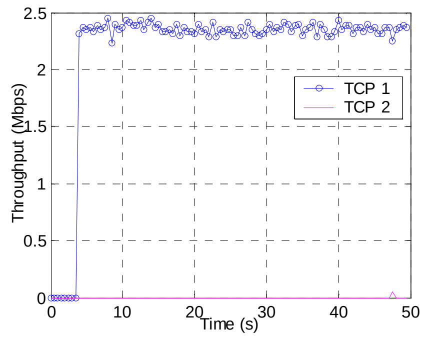
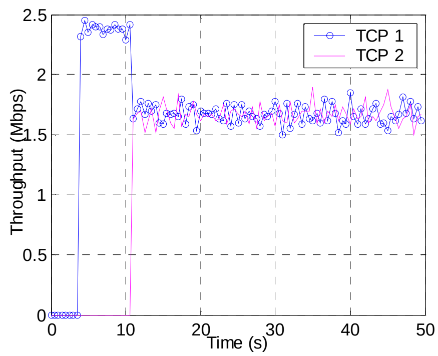


Fig. 8. Topologies and traffic flows being simulated for study of HN effects.

"TCP unfairness" and "re-routing instability" are two performance problems triggered by HN identified previ-



(a) Without HFD



(b) With HFD

Fig. 9. TCP fairness (unfairness) with (without) HFD.

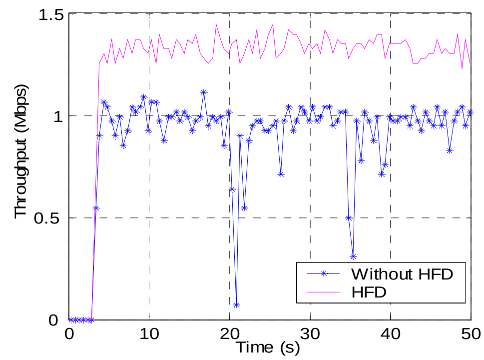


Fig. 10. Throughput stability (instability) of an 11-hop UDP flow with (without) HFD.

ously [5]. This section validates by NS-2 simulation that HFD eliminates these problems. As in [5], we consider a chain topology, as shown in Fig. 8. In (a), 6 nodes in a straight line are spaced apart equally by 140 meters; in (b), 12 nodes in a straight line are spaced apart equally by 140 meters. The data rate is set at 11Mbps. The two-ray ground propagation model is adopted with loss exponent $\alpha = 4$. The thresholds C_{rt} and C_t, are both set to 10. The RS mode is turned on for the HFD simulation and turned off the non-HFD simulation. The physical carrier-sensing range is 550m in both cases. Thus, the d_{max} for HFD according to (24) is 550/3.78 = 145m, and both the topologies (a) and (b) satisfy this. The Ad-hoc On-Demand Distance Vector (AODV) routing protocol [21] is used, and TCP/UDP packet size of 1460 Bytes is assumed.

8.1 TCP Unfairness

We ran a TCP simulation experiment on the topology of Fig. 8 (a). TCP 1 is from node 1 to node 3, and TCP 2 is from node 6 to node 4. TCP 1 starts earlier than TCP 2 at time = 3.0 sec, and TCP 2 starts at time = 10.0 sec. Without HFD's RS mode, node 1 is hidden from node 5: node 4 can sense node 1 but not node 5, causing node 5's DATA packet to be ignored by node 4 when node 1's DATA is already in progress. Likewise, node 2 is hidden from node 6. Because TCP 1 starts earlier, TCP 2 virtually has no chance to obtain any throughput (See Fig. 9 (a)). Fig. 9 (b) shows that this severe "unfairness" problem is eliminated with HFD.

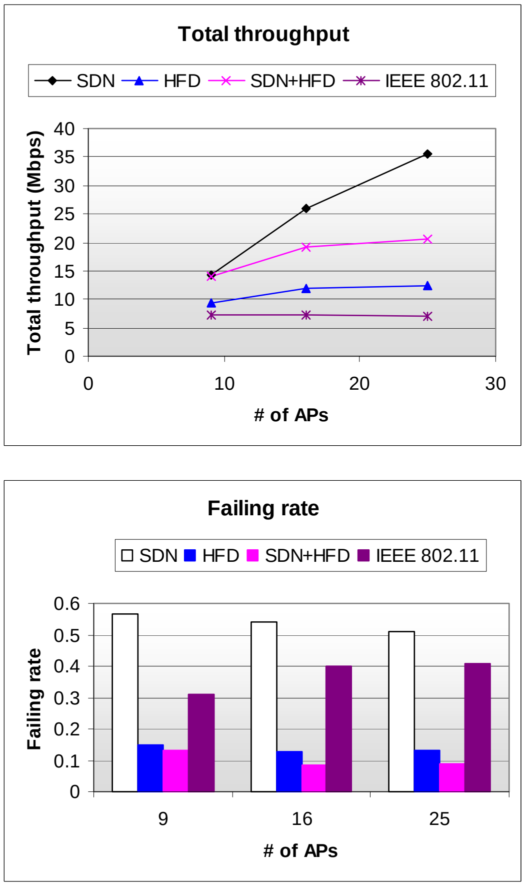
8.2 Re-routing Instability

We performed a UDP simulation experiment on the topology of Fig. 8 (b). There is a UDP flow from node 1 to node 12. Without HFD's RS mode, node 5 is hidden from node 1, causing the DATA packets of node 1 to be repeatedly ignored at node 2 when node 5 transmits its DATA packets. Likewise, nodes 2, 3, 4, … face the same problems. It has been shown in [5] [9] that throughput instability can result from misinterpretation by the routing algorithm that links are down because of such repetitive packet transmission failures. Reference [9] showed that this instability can be removed by de-activating the part of the routing algorithm that gives up the old route before a new route can be found. Fig. 10 shows that the throughput instability can also be removed by HFD.

9 Combination of SDN and HFD

TABLE 4
Combination of SDN and HFD for RTS/CTS Access Mode

	Receiver Restart Mode	Power Exchange Algorithm	Selective Disregard of NAV	Power Budget Requirements of HFD	PCS for RTS/CTS
SDN	x	x	x		
HFD	x			x	x
SDN+HFD	x	x	x	x	x



Failing rate = # of failed transmissions / total # of transmissions

Fig. 11. Throughput Comparison (closely related to EN, and loosely related to HN) and Failing Rate Comparison (related to HN).

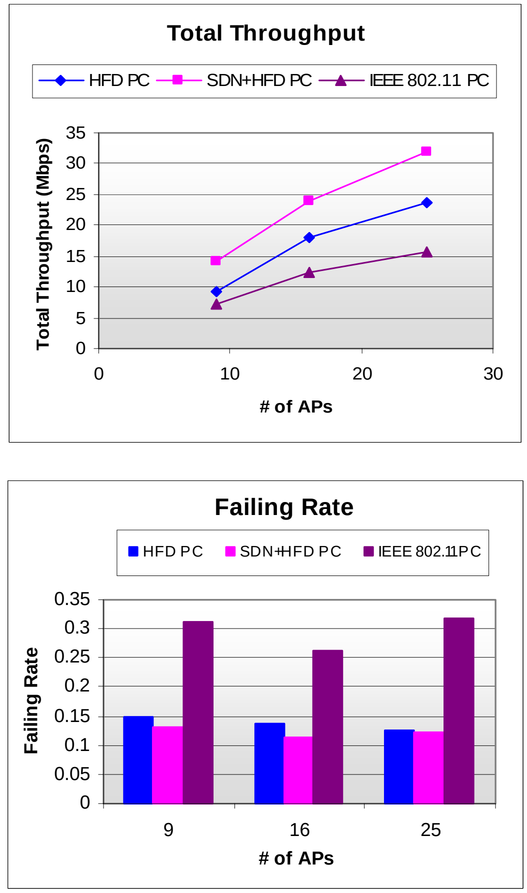


Fig. 12. Comparison of Throughput and Failing Rate (With Power Control—"PC").

We have considered SDN and HFD separately in the last four sections. We now combine them in the fashion shown in Table 4. Note that the "HFD" refers to "HFD for RTS/CTS Access Mode" here, since SDN works in this mode can only be combined with this version of HFD. To demonstrate SDN+HFD can mitigate EN and HN simultaneously, we carried out NS-2 simulations with the grid topology as in Fig. 7 (a), where D is set to 700m and M = 3, 4, 5. A total of $4M^2$ clients are randomly placed in the whole area with a uniform distribution. Each client connects to the nearest AP at the center of a cell. Therefore, $d_{\max} = D/(\sqrt{2}M)$ (the distance from the center to a corner of one cell). We set C_t=C_{rt}=10, PCS=640m, VCS=480m. With the above settings, it can be verified that the two power budget requirements of HFD, (31) and (32), are satisfied.

9.1 No power control

We first consider the case in which there is no power control so that as M increases, the PCS and VCS ranges remain 640m and 480m respectively. From Fig. 11, pure SDN has the best throughput performance. SDN eliminates EN entirely, which enabling scalable throughput as number of AP increases, despite having no power control. However, the higher degree of HN in SDN is revealed in the transmission failure rate due to HN. (Despite the failing rate, SDN has the highest throughput because the number of transmission attempts is the highest.)

The throughputs of the original 802.11, HFD, and SDN+HFD do not scale. The reason that HFD and SDN+HFD do not scalable throughput is because physical carrier sensing is turned on for RTS/CTS in HFD. When the coverage area of a fixed PCS range is already heavily populated with simultaneous RTS/CTS, further adding APs does not help spatial re-use due to EN of RTS/CTS.

It is interesting to note that both HFD and SDN+HFD can achieve higher throughput and lower transmission failure rate than the original 802.11. For HFD, the throughput increase is due to the lower failing rate, which results from reduced HN. *For SDN+HFD, this can be explained by the fact that both EN and HN are reduced compared to 802.11*, where both EN and HN are severe.

Note that even with HFD, the failing rate is about 10%. Most of these failings are not due to HN, but due to the fact that different nodes may still happen to transmit at the same time slot (i.e., their random backoff counter reduces to 0 at the same time). This is a different kind of collisions which do not cause much trouble because after the collision, new random backoff numbers are picked by these nodes and consecutive collisions are unlikely.

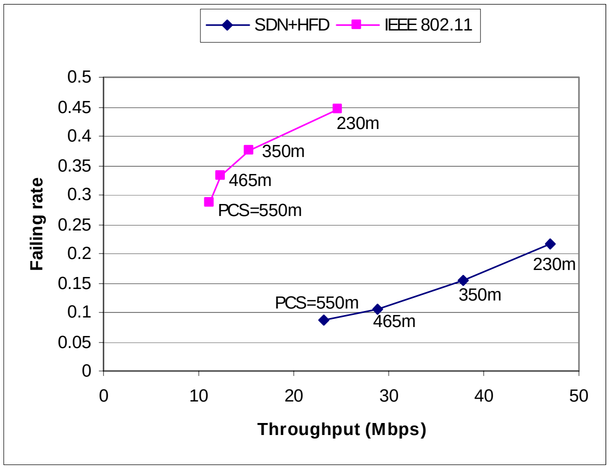


Fig. 13. Tradeoff between EN and HN (Total Throughput vs. Failing rate. The label besides each point indicates the PCS range used at that point).

9.2 Uniform power control (or receiving threshold adjustments)

In the next experiment, we scaled *PCS* and *VCS* down in proportion to the cell size as *M* increases. For the 16-AP case, the ranges are scaled by a factor of 3/4; and for the 25-AP case, by a factor of 3/5. This can be implemented by two alternative power/threshold adjustment schemes, as discussed below. Since d_{max} is also scaled down proportionally, the power budget requirements of HFD are still satisfied. If two-ray ground propagation model with $\alpha = 4$ is assumed, (25)-(27) can written as

$$P_{threshold} = P(P_t; Range) \propto P_t / Range^4 \tag{33}$$

where $P_{threshold}$ stands for any of P_{link}, P_{PCS} and P_{VCS}, and *Range* stands for any of d_{max}, *PCS* and *VCS*, respectively.

Scheme 1: *Adjust* P_t *with fixed* P_{link}, P_{PCS} *and* P_{VCS}

According to (33), to reduce *Range* to a fraction of *G* times the original value, we can reduce P_t of all the nodes to a fraction of G^4 times the original level, while keeping the threshold values P_{link}, P_{PCS}, P_{VCS} unchanged. Thus, for G=3/4, G^4=0.316; for G=3/5, G^4=0.130.

Scheme 2: *Adjust* P_{link}, P_{PCS} *and* P_{VCS} *with fixed* P_t

Fix P_t but raise $P_{threshold}$ (P_{link}, P_{PCS} and P_{VCS}) of all the nodes to G^{-4} times the original values. For G=3/4, G^{-4}=3.16; for G=3/5, G^{-4}=7.69.

As can be seen from Fig. 12, with uniform power control ("*Scheme 1*" is adopted in our simulation. But "*Scheme 2*" should give the same results), all schemes become scalable. And HFD and SDN+HFD still have better throughput performance than the original 802.11. Note that the throughput performance of pure SDN is the same with or without power control and is therefore not drawn here. SDN+HFD is superior to HFD and 802.11 in terms of both network throughput and transmission failure rate.

9.3 Non-uniform power control or receiving threshold adjustment

Here, different nodes can use different transmit powers or receiving thresholds, which can be set properly after a node finds out its interference relationship with the neighboring links through the Power-Exchange Algorithm. This is a fully distributed algorithm, and it is especially useful in networks where the nodes are not uniformly distributed. In [20], we have explored this scheme for IEEE 802.11 Basic Access Mode. As part of future work, it would be interesting to extend it to the RTS/CTS Access Mode to be combined with SDN.

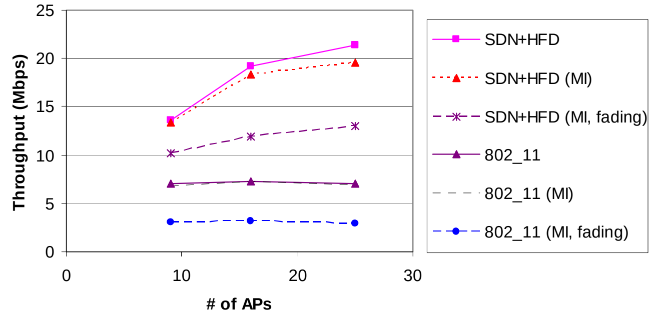


Fig. 14. The effect of Multiple Interference ("MI") and Channel Fading.

9.4 Tradeoff between EN and HN

In the above, we have seen that EN and HN can not be simultaneously eliminated by SDN+HFD, especially in the case without power control. To reduce the collisions (HN) among RTS/CTS packets, PCS for RTS/CTS is turned on. This, however, introduces EN among RTS/CTS packets. Therefore, there is a fundamental tradeoff between EN and HN in SDN+HFD (the tradeoff certainly exists in the original 802.11). If we decrease the PCS range of RTS/CTS, EN is reduced but HN is increased.

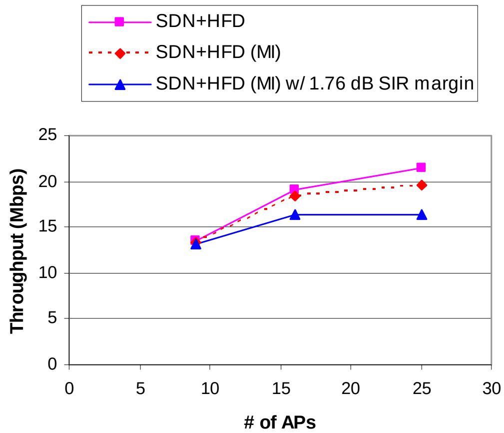
To illustrate this point, we vary the PCS range (and also VCS range proportionally) in the above 16-AP network. The total throughput and failing rate as functions of PCS range are plotted in Fig. 13 (the label besides each point indicates the PCS range at that point). For comparison, the curves of 802.11 are also plotted.

However, it is also important to note that with SDN+HFD, both EN and HN have been simultaneously reduced compared to 802.11, which leads to higher total throughput and lower failing rate. In summary, although the fundamental tradeoff between EN and HN still exists, "tradeoff frontier" becomes much better than that in the original 802.11 (see Fig. 13).

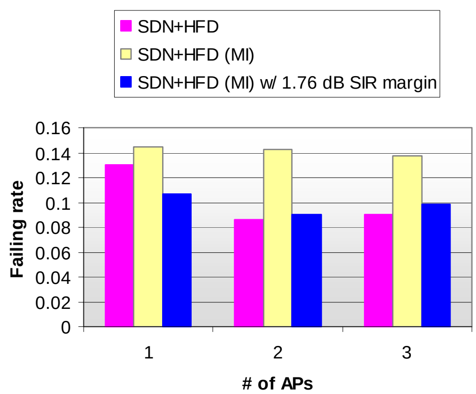
9.5 Effect of Multiple Interference ("MI") and Channel Fading

By now, we have designed and evaluated several protocols based on a mathematical model to address EN and HN in 802.11 wireless networks. In practice, however, there are some other issues to consider. For example, interference at a receiver may come simultaneously from the multiple nodes, whose powers accumulate at the receiver and further reduce the SIR. Also, channel fadings make the received power random instead of deterministic. And the fluctuation of received power may introduce EN or HN in our design.

In this subsection, we consider the effects of multiple



(a) Total Throughput



(b) Failing Rate

Fig. 15. Using SIR margin to reduce packet collisions.

interference and channel fading on the performance of SDN+HFD, as well as the original IEEE 802.11. And we propose to use a "perturbation design" (i.e., introduce a "SIR margin") to counter some of the negative effects.

We have modified the original NS-2 code to take into account the above effects. For each protocol (SDN+HFD or 802.11, w/o power control), we simulate the case when multiple interference ("MI") is present, and the case when both MI and fading are present. We assume Rayleigh fading, where the received power is multiplied by a factor which is exponentially distributed with expected value 1. (Rayleigh fading assumes no "Line-of-Sight" propagation. In practice if there is a direct "Line-of-Sight" propagation path, the performance would be much better.) Fig. 14 shows the results of "SDN+HFD" and "802.11" (other protocols have similar patterns).

We have the following observations:

- The performance of both protocols suffers from multiple interference and fading.
- The effect of multiple interference is not significant in both protocols. Therefore, the pairwise interference model is a good approximation, with the advantage of simplifying protocol designs. In Fig. 14, the effect of MI on 802.11 seems smaller, since carrier-sensing in 802.11 is quite conservative such that the accumulation of multiple interference does not make the SIR level drop below the threshold (i.e., 10) very often.
- Rayleigh fading has a larger effect. With the received power exponentially distributed, the probability of very low power is considerable. In Fig. 14, it actually cuts down a larger percentage of the throughput from 802.11 than SDN+HFD.

We can introduce a "SIR margin" to offset the extra collisions caused by MI. That is, instead of implementing SDN+HFD with SIR threshold 10, we target a higher SIR threshold. With this change, the PCS/VCS ranges need to be increased, and there are more "Interference Relationships" assumed by each link. Fig. 15 shows the simulation results before and after setting SIR threshold to 15 (i.e., 11.76dB, with a 1.76dB-margin above 10dB).

With this SIR margin, the "failing rate" roughly returns to the level without considering MI (meaning that the extra collisions caused by MI are cancelled out), but as a tradeoff, the throughput is decreased due to the increased PCS/VCS ranges.

SIR margin can also be used to combat channel fading. But it is more difficult to address due to the possibility of very low received power (deep fading). However, the performance degradation caused by fading is a common problem in all wireless networks, and not a problem specific to our design (Fig. 14).

10 Conclusion

In the seminal work of [1] [2], the authors established the fundamental capacity bounds for large-scale wireless networks. This work led to numerous follow-up investigations by the research community. Most of these efforts were theoretical in nature and they generally assumed “perfect scheduling” in which there is implicitly an all-knowing scheduler with global information to schedule node transmissions. Meanwhile, in the practical realm, the IEEE 802.11 WLAN based on a CSMA/CA MAC protocol has enjoyed tremendous commercial success. There is a gap between theory and practice today.

This paper is an attempt to fill the gap. Although there have been many investigations on the well-known exposed-node (EN) and hidden-node (HN) phenomena in 802.11 networks, most of these investigations are based on incidental examples. This paper has provided formal definitions for EN and HN needed for a more comprehensive study. We have shown that EN leads to non-scalable throughput as node density increases. In particular, non-scalability is fundamentally due to EN rather than HN. HN, on the other hand, causes excessive collisions, unfair bandwidth distributions and in the extreme case, bandwidth starvation.

We have devised a variant of the 802.11 MAC protocol called Selective Disregard of NAVs (SDN) which achieves network throughput scalability by removing EN entirely. In particular, we have argued that the O(n) scalability based on perfect scheduling [2] is also achievable with CSMA/CA scheduling in infrastructure networks. In addition, we have derived a set of criteria, called Hidden-Node Free Design (HFD) to overcome HN. This is in contrast to many previous efforts which attempt to deal with the performance problems created by HN rather than get to the root of the problem to remove HN directly. We have shown that HFD can be readily combined with SDN to achieve the advantages of both. In particular, the SDN+HFD, together with power control, can yield scalable and significantly higher throughput than the original 802.11, while minimizing packet collisions due to HN.

In Section 9, we have shown that there is a fundamental tradeoff between EN and HN, in both IEEE 802.11 and the proposed "SDN+HFD". However, "SDN+HFD" gives a better “tradeoff frontier” by reducing EN and HN simultaneously. An interesting direction for future research is a more analytical study of this tradeoff in a given network (deterministically), or in a random network (probabilistically), with properly defined measures of EN and HN, such as 1) The number of EN-causing and HN-causing edges; 2) Total throughput (reflecting EN) and failing rate (reflecting HN).

Appendix I: Condition for Correct Operation of PE (Power-Exchange Algorithm)

The following condition is sufficient to ensure a node can discover all its relevant s-edges:

$$P(PERange-d_{\max}) < P(d_{\max})/K \qquad (34)$$

where *PERange* is the transmission range of the *PE packets*, and *P*(.) is the received power, which is assumed to be a decreasing function of distance.

Comment: To meet (34), PE packets can be transmitted at a sufficiently low rate (e.g., the same rate as RTS/CTS packets).

Proof: Consider three nodes: nodes a and b form links (a, b) and (b, a); and node c form links with other nodes. Without loss of generality, assume that P(a,b) < K P(c,b) so that the transmission by c can interfere with reception at b. We want to show that (i) nodes a and b, and (ii) node c and any other node d that forms link with c, can find out the existence of the associated s-edges.

Proof of (i): By definition, we have $|a-b| \le d_{\max}$. If (34) holds, we have

$$K\,P(c,b) > P(a,b) \ge P(d_{\max}) > K\,P(PERange-d_{\max}) \qquad (35)$$

The above implies *PERange* - d_{max} > $|c - b|$, since *P*(.) is a decreasing function of distance. So,

$$\begin{aligned} &PERange > |c-b|, \text{ and} \\ &PERange > |c-b| + d_{\max} \ge |c-b| + |a-b| \ge |c-a| \end{aligned} \qquad (36)$$

This means that if P(a,b) < K P(c,b), then the PE packets of c can reach a and b. By measuring the received power of c’s PE packets and checking their source address to identify the sender, b can derive c and P(c, b). Similarly, b can derive a and P(a, b) from a’s PE packets. Therefore, b can find out that P(c, b) > K P(a, b), and hence the existence of s-edges between link (a, b)/(b, a) and any other link with c being the transmitter or receiver – note that these links are contained in the active link list in c’s PE packets received by b.

Now, the PE packets of b contains information on [c, P(c, b)] and [a, P(a, b)]. Upon receiving b’s PE packets, node a can also find out P(c, b) > K P(a, b), and hence the existence of s-edges between link (a, b)/(b, a) and any other link with c being the transmitter or receiver – note that these links are contained in the active link list in c’s PE packets received by a.

Proof of (ii): From the proof of (i), we have *PERange*-d_{max} > $|c\text{-}b|$. So,

$$\begin{aligned} &PERange > |b-c|, \text{ and} \\ &PERange > |b-c| + d_{\max} \ge |b-c| + |c-d| \ge |b-d| \end{aligned} \qquad (37)$$

Thus, the PE packets of b, which contain [c, P(c, b)], [a, P(a, b)] and active links (a, b) and (b, a), can reach c and any node d that forms link with c, from which they can establish the associated s-edges. ☐

Appendix II: Comparison of the Scalability of SDN and 802.11

II.A Scalability of SDN

This and the next subsections give the analytical arguments for the observations in Section 6. We first argue for the scalability of SDN. Consider a very large-scale wireless network with random placement of client nodes with certain density distribution across an infinitely large area *A*. For a given client-AP ratio, assume the APs are placed regularly in a grid manner across *A*. Let *n* be the number of client nodes per unit area. Then, the average throughput per unit area is proportional to *n* for a given client-AP ratio. The corresponding average transport capacity is proportional to $\sqrt{n}$. To see this, given an “original” node distribution in *A*, suppose we map the position of node *i* from X_i to $X'_i = k*X_i$, where $k < 1$. An area *a* in *A* is then mapped to an area of *a’* in *A* in the transformed domain.

This scaling does not affect the inequalities (1) - (8) (more exactly, the equivalent distance relationships given by the power relationships of (1) – (8) [13]). That is, two links have interference relationships in the transformed domain if and only if they have interference relationships in the original domain. To the extent that spatial reuse is completely characterized by (1) – (8), which is the case in SDN, the throughput of the nodes within *a’* is the same as the throughput within *a*.

In the transformed domain, however, the average throughput within *a* has increased by a factor of $1/k^2$. The average number of nodes in *a* also increases by a factor of $1/k^2$. Therefore, the average throughput per node remains the same after the transformation. Thus, the average throughput per unit area is of order *n* where *n* is the number of client nodes per unit area. The average distance between a client node and its associated AP is reduced by a factor of *k* in the transformed domain. So, the transport capacity [1] is proportional to $\sqrt{n}$. The whole problem is akin to transforming a map to a smaller-scale map – except for the drawing scale, nothing has changed.

II.B Non-Scalability of 802.11

A similar “transformation” argument shows that the average throughput in 802.11 *does not* scale with *n*. The reason is that some spatial-reuse inequalities characterizing 802.11 – from (9) to (15) – are not invariant under the transformation (if the transmit powers are not scaled down at the same time). In particular, increasing number of APs and client node density does not change the fact that each transmission uses the same amount of "spatial resources". Thus, average throughput per unit area remains constant as node density increases. The average transport capacity is proportional to $1/\sqrt{n}$ assuming each client node associates with the closest AP.

Even in an "arbitrary network" (as defined in [1]), where one has the freedom to control the nodes' positions and traffic flows as one wishes, it can be proved that the "best" transport capacity and overall throughput do not exceed the order of $O(1)$. In other words, they do not scale with the increase of n. The detailed proof, given in [19], applies the concept of "exclusion region" introduced in [1] on 802.11. To conserve space, the proof will not be repeated here.

Acknowledgment

This work was partially supported by the Competitive Earmarked Research Grant (Project Number 414106) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

References

[1] P. Gupta, P. R. Kumar, "The Capacity of Wireless Network", *IEEE Trans. on Information Theory*, vol. 46, no. 2, pp.388-404, Mar. 2000.
[2] Ashish Agarwal and P. R. Kumar, "Capacity bounds for ad-hoc and hybrid wireless networks." *ACM SIGCOMM Computer Communications Review, Special Issue on Science of Networking Design*, vol. 34, no. 3, pp. 71-81, July 2004.
[3] IEEE Standards Department, *IEEE 802.11 Standard for Wireless LAN, Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1999.
[4] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LANs," *Proc. ACM SIGCOMM*, vol. 24, Oct. 1994, pp. 212–225.
[5] Xu, S.; Saadawi, T., "Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks?" *IEEE Communications Magazine*, vol. 39, no. 6, pp. 130 – 137, Jun. 2001.
[6] X. L. Huang, B. Bensaou, "On Max-min Fairness and Scheduling in Wireless Ad-Hoc Networks: Analytical Framework and Implementation", *ACM MobiHoc'01*, Oct. 2001, pp. 221-231.
[7] S. Ansari et al. "Performance Enhancement of TCP on Multihop Ad hoc Wireless Networks", *IEEE ICPWC'02*, Dec. 2002, pp. 90-94.
[8] P. C. Ng, S. C. Liew, "Offered Load Control in IEEE 802.11 Multi-hop Ad-hoc Networks," the *1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, Oct. 2004.
[9] P. C. Ng, S. C. Liew, "Re-routing Instability in IEEE 802.11 Multi-hop Ad-hoc Networks, " the *4th IEEE International Workshop on Wireless Local Network* , Nov. 2004.
[10] Monks, J.P., Bharghavan, V., Hwu, W.-M.W., "A power controlled multiple access protocol for wireless packet networks," *Proc. IEEE INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies,* vol. 1, Apr. 2001, pp. 219 - 228.
[11] A. Muqattash, M. Krunz, "Power controlled dual channel (PCDC) medium access protocol for wireless ad hoc networks," *Proc. IEEE INFOCOM*, 2003, pp. 470-480.
[12] Haas, Z.J., Jing Deng, "Dual busy tone multiple access (DBTMA)-a multiple access control scheme for ad hoc networks Communications," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 975 – 985, Jun. 2002.
[13] P. C. Ng, S. C. Liew, and L. B. Jiang, "Achieving Scalable Performance in Large-Scale IEEE 802.11 Wireless Networks," *IEEE Wireless Communications & Networking Conference*, Mar. 2005.
[14] "The Network Simulator: ns2", http://www.isi.edu/nsnam/ns.
[15] T. Rappaport, "Wireless Communications: Principles and Practice," Prentice Hall, New Jersey, 2002.
[16] S. McCanne, S. Floyd, The ns Manual, Obtain via: http://www.isi.edu/nsnam/ns/doc/ns_doc.pdf.
[17] K. Xu, M. Gerla, S. Bae, "How Effective is the IEEE 802.11 RTS/CTS Handshake in Ad Hoc Networks?" *IEEE GLOBECOM'02*, vol. 1, Nov. 2002, pp. 17-21.
[18] S. Bansal, R. Shorey, A. A. Kherani, "Performance of TCP and UDP Protocols in Multi-Hop Multi-Rate Wireless Networks", *IEEE Wireless Communications & Networking Conference*, Apr. 2004.
[19] L. B. Jiang, "Improving Capacity and Fairness by Elimination of Exposed and Hidden Nodes in 802.11 Networks," *MPhil Thesis*, the Chinese University of Hong Kong, Jun. 2005. (http://www.eecs.berkeley.edu/~ljiang/Thesis.pdf)
[20] L. B. Jiang, S. C. Liew, "Removing Hidden Node Problem in IEEE 802.11 Wireless Networks," *IEEE Vehicular Technology Conference, Fall 2005.*
[21] C. E. Perkins, E. M. Royer, "Ad-hoc On-Demand Distance Vector Routing," *Proc. 2nd IEEE workshop on Mobile Computing Systems and Applications*, Feb. 1999, pp. 90-100.

Li Bin Jiang received his B.Eng. degree in Electronic Engineering & Information Science from the University of Science and Technology of China in 2003 and the M.Phil. degree in Information Engineering from the Chinese University of Hong Kong in 2005, and is currently working toward the Ph.D. degree in the Department of Electrical Engineering & Computer Science, University of California, Berkeley. His research interest includes wireless networks, communications and computer networks.

Soung Chang Liew (S'87–M'88–SM'92) received his S.B., S.M., E.E., and Ph.D. degrees from the Massachusetts Institute of Technology. From 1984 to 1988, he was at the MIT Laboratory for Information and Decision Systems, where he investigated Fiber-Optic Communications Networks. From March 1988 to July 1993, Soung was at Bellcore (now Telcordia), New Jersey, where he engaged in Broadband Network Research. He is currently Professor and Chairman of the Department of Information Engineering, the Chinese University of Hong Kong. Soung's current research interests include wireless networks, Internet protocols, multimedia communications, and packet switch design. Recently, Soung and his student won the best paper awards in the 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2004) the 4th 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 Soung and his student, has been incorporated into a recent release of Linux OS. Besides academic activities, Soung is also active in the industry. He co-founded two technology start-ups in Internet Software and has been serving as consultant to many companies and industrial organizations. He is currently 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. Soung is the holder of three U.S. patents and Fellow of 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. Publications of Soung can be found in www.ie.cuhk.edu.hk/soung.