Performance of Collision Avoidance Protocols in Single-Channel Ad Hoc Networks *

Yu Wang J.J. Garcia-Luna-Aceves
Department of Computer Engineering
University of California
Santa Cruz, CA 95064, U.S.A.
Email: {ywang, jj}@cse.ucsc.edu
URL: http://www.cse.ucsc.edu/research/ccrg

Abstract

This paper presents the first analytical model to derive the saturation throughput of collision avoidance protocols in multi-hop ad hoc networks with nodes randomly placed according to a two-dimensional Poisson distribution. We show that the sender-initiated collision-avoidance scheme performs much better than the ideal CSMA scheme with a separate channel for acknowledgments. More importantly, we show that the collision-avoidance scheme can accommodate much fewer competing nodes within a region in a network infested with hidden terminals than in those cases without hidden terminals or with just a few, if reasonable throughput is to be maintained. Simulations of the popular IEEE 802.11 MAC protocol show that it cannot ensure collision-free transmission of data packets and thus throughput can degrade well below what is predicted by the analysis of a correct collision avoidance protocol. Based on these results, a number of improvements are proposed for the IEEE 802.11 MAC protocol.

Keywords

Collision avoidance, MAC, ad hoc networks, fairness, IEEE 802.11, analytical modeling, simulation evaluation,

1 Introduction

Because of the "hidden terminal" problem, the performance of simple MAC protocols like the carrier sense multiple access (CSMA) protocol degrades to that of the ALOHA protocol in ad hoc networks. Thus, many collision-avoidance protocols (e.g., [1, 2]) have been proposed in the recent past to combat the hidden terminal problem. The most popular collision avoidance scheme today consists of a sender-initiated four-way handshake in which the transmission of a data packet and its acknowledgment is preceded by request-to-send (RTS) and clear-to-send (CTS) packets between a pair of sending and receiving nodes. Other nodes that overhear RTS or CTS packets will defer their access to the channel to avoid collisions.

Although a salient feature of ad hoc networks is that, for any node in such networks, potential interference from hidden nodes always exists, the analytical models of collision-avoidance protocols are largely confined to single-hop networks [3–5] or cases when the number of hidden terminals is very small [6]. Hence, we are interested in investigating the performance of the sender-initiated collision avoidance scheme based on a four-way handshake in a more general framework applicable to ad hoc networks. We focus on the case of single-channel ad hoc networks in which omnidirectional antennas are used.

In this paper, we adopt a simple model to derive the saturation throughput of this sender-initiated collision avoidance scheme. In this network model, nodes are randomly placed on a plane according to two-dimensional Poisson distribution with density λ. In this model, it is also assumed that each node is ready to transmit independently in each time slot with probability p, where p is a protocol-dependent parameter. This model was first used by Takagi and Kleinrock [7] to derive the optimum transmission range of a node in a multi-hop wireless network, and was used subsequently by Wu and Varshney [8] to derive the throughputs of non-persistent CSMA and some variants of busy tone multiple access (BTMA) protocols.

In Section 2, we use the model to analyze the sender-initiated collision-avoidance scheme based on a four-way handshake and non-persistent carrier sensing. To our knowledge, this is the first analytical model of collision avoidance in multi-hop networks. To simplify our task, we first assume that both carrier sensing and collision avoid-

*This work was supported in part by the US Air Force/OSR under Grant No. F49620-00-1-0330.

ance work perfectly, that is, that nodes can accurately sense the channel busy or idle, and that the RTS/CTS scheme can avoid the transmission of data packets that collide with other packets at the receivers. The latter assumption can be called *perfect collision avoidance* and has been shown to be doable in the floor acquisition multiple access (FAMA) protocol [2]. We then extend this model to take into account the possibility of data packets colliding with other transmissions.

In Section 3, we present numerical results from our analysis. We compare the performance of the sender-initiated collision avoidance scheme against an idealized non-persistent CSMA protocol. It is shown that the RTS/CTS scheme can achieve far better throughput than the CSMA protocol, even when the overhead due to RTS/CTS exchange is high. The results illustrate the importance of enforcing collision avoidance in the RTS/CTS handshake. More importantly, our results show that hidden terminals degrade the performance of collision avoidance protocols beyond the basic effect of having a longer vulnerability period for RTSs [2]. Hence, it follows that collision avoidance becomes more and more ineffective for a relatively crowded region with hidden terminals.

To validate the findings drawn from this analysis, in Section 4 we present simulations of the popular IEEE 802.11 MAC protocol. The simulation results clearly show that the IEEE 802.11 MAC protocol cannot ensure collision-free transmission of data packets, and that almost half of the data packets transmitted cannot be acknowledged due to collisions. We also investigate a variant of the IEEE 802.11 MAC protocol in which the contention window used in deciding backoff time is fixed. This variant does not have the inherent fairness problem in the original backoff scheme used in the IEEE 802.11 MAC protocol, though it is not fine tuned to achieve the best performance. However, the simulation results do show that decreasing the contention window leads to more collisions of data packets, while increasing the contention window leads to more wasted time in waiting. The performance of the simulated IEEE 802.11 correlates well with what is predicted in the extended analysis, which takes into account the effect of data packet collisions and is used for the case when the number of competing nodes in a region is small. The simulation results for IEEE 802.11 show a larger variation in throughput than the predicted performance from the analytical model, which is possibly due to its inherent fairness problems. When the number of competing nodes in a region increases, the performance gap between IEEE 802.11 and the analysis decreases, which validates the statement that even a perfect collision-avoidance protocol loses its effectiveness gradually. Section 5 concludes this paper with possible ways to improve the performance of collision avoidance protocols in ad hoc networks.

2 Approximate Analysis

According to our network model, the nodes are Poisson distributed over a plane with density λ, i.e., the probability $p(i,S)$ of finding i nodes in an area of S is given by:

$$p(i,S) = \frac{(\lambda S)^i}{i!} e^{-\lambda S}.$$

Assume that each node has the same transmission and receiving range of R, and denote by N the average number of nodes within a circular region of radius R; therefore, we have $N = \lambda \pi R^2$.

To simplify our analysis, we assume that nodes operate in time-slotted mode. As prior results for CSMA and collision-avoidance protocols show, the performance of MAC protocols based on carrier sensing is much the same as the performance of their time-slotted counterparts in which the length of a time slot equals one propagation delay and the propagation delay is much smaller than the transmission time of data packets.

The length of each time slot is denoted by τ. Note that τ is not just the propagation delay, because it also includes the overhead due to the transmit-to-receive turn-around time, carrier sensing delay and processing time. In effect, τ represents the time required for all the nodes within the transmission range of a node to know the event that occurred τ seconds ago. The transmission times of RTS, CTS, data, and ACK packets are normalized with regard to τ, and are denoted by l_{rts}, l_{cts}, l_{data}, and l_{ack}, respectively. Thus, τ simply equals 1. For simplicity, we also assume that all packet lengths are multiples of a time-slot length.

We derive the protocol's throughput based on the heavy-traffic assumption, i.e., a node always has a packet in its buffer to be sent and the destination is chosen randomly from one of its neighbors. This is a fair assumption in ad hoc networks in which nodes are sending data and signaling packets continually. We also assume that a silent node may be ready with probability p at each time slot. Here p is a protocol-specific parameter but is slot independent, which is also assumed in previous work [5,7,8]. Even when the node is ready to transmit, it may transmit or not in the slot, depending on the collision avoidance and resolution schemes being used, as well as the channel's current state. Thus, we are more interested in the probability that a node transmits in a time slot, which is denoted by p'. Like Takagi and Kleinrock [7], we also assume that p' is independent at any time slot to make the analysis tractable. Given this simplification, p' can be defined to be:

$$\begin{aligned} p' &= p \cdot \text{Prob.}\{\text{Channel is sensed idle in a slot}\} \\ &\approx p \cdot \Pi_I \end{aligned}$$

where Π_I is the limiting probability that the channel is in idle state, which we derive subsequently.

We are not interested in the exact relationship between p and p', and it is enough to obtain the range of values that

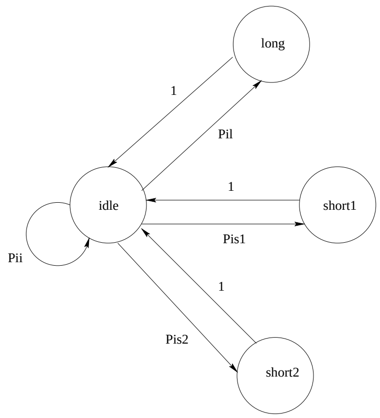


Figure 1. Markov chain model for the channel around a node

p' can take, because the throughput of these protocols is mostly influenced by p'. The purpose of setting up the channel model is mostly to derive the rough relationship between p and p'. To simplify the channel model, we make two key assumptions: First, we model the channel as a circular region in which there are some nodes. The nodes within the region can communicate with each other while they have weak interactions with nodes outside the region. *Weak interaction* means that the decision of inner nodes to transmit, defer and back off is almost not affected by that of outer nodes and vice versa. Considering that nodes do not exchange status information explicitly (e.g. either defer due to collision avoidance or back off due to collision resolution), this assumption is reasonable and helps to simplify the model considerably. Thus the channel's status is only decided by the successful and failed transmissions within the region.

Second, we still consider the failed handshakes initiated by nodes within the region to outside nodes, because this has a direct effect on the channel's usability for other nodes within the region. Though the radius of the circular region R' is unknown, it falls between $R/2$ and $2R$. This follows from noting that the maximal radius of a circular region in which all nodes are guaranteed to hear one another equals $R' = R/2$, and all the direct neighbors and hidden nodes are included into the region when $R' = 2R$. Thus, we can write $R' = \alpha R$ where $0.5 \leq \alpha \leq 2$, and α needs to be estimated.

With the above assumptions, the channel can be modeled by a four-state Markov chain illustrated in Figure 1. The significance of the states of this Markov chain is the following: *Idle* is the state when the channel around node x is sensed idle, and obviously its duration is τ. *Long* is the state when a successful four-way handshake is done. For simplicity, we assume that the channel is in effect busy for the duration of the whole handshake, thus the busy time T_{long} is

$$\begin{aligned} T_{long} &= l_{rts} + \tau + l_{cts} + \tau + l_{data} + \tau + l_{ack} + \tau \\ &= l_{rts} + l_{cts} + l_{data} + l_{ack} + 4\tau. \end{aligned}$$

Short1 is the state when multiple nodes around the channel transmit RTS packets during the same time slot and their transmissions collide. The busy time of the channel T_{short1} is therefore $T_{short1} = l_{rts} + \tau$. *Short2* is the state when one node around the channel initiates a failed handshake with a node outside the region. Even though a CTS packet may not be sent due to the collision of the sending node's RTS packet with other packets originated from nodes outside the region or due to the deferring of the receiving node to other nodes, those nodes overhearing the RTS as well as the sending node do not know if the handshake is successfully continued, until the time required for receiving a CTS packet elapses. Therefore the channel is in effect busy, i.e., unusable for all the nodes sharing the channel, for the time stated below:

$$\begin{aligned} T_{short2} &= l_{rts} + \tau + l_{cts} + \tau \\ &= l_{rts} + l_{cts} + 2\tau. \end{aligned}$$

Now we proceed to calculate the transition probabilities of the Markov chain.

In usual collision avoidance schemes, no node is allowed to transmit immediately after the channel becomes idle, thus the transition probabilities from *long* to *idle*, from *short1* to *idle* and from *short2* to *idle* are all 1.

According to the Poisson distribution of the nodes, the probability of having i nodes within the receiving range R of x is $e^{-N}N^i/i!$, where $N = \lambda\pi R^2$. Therefore, the mean number of nodes that belong to the shared channel is $M = \lambda\pi R'^2 = \alpha^2 N$. Assuming that each node transmits independently, the probability that none of them transmits is $(1-p')^i$ where $(1-p')$ is the probability that a node does not transmit in a time slot. Because the transition probability P_{ii} from *idle* to *idle* is the probability that none of the neighboring nodes of x transmits in this slot, P_{ii} is given by

$$\begin{aligned} P_{ii} &= \sum_{i=0}^{\infty}(1-p')^i \frac{M^i}{i!} e^{-M} \\ &= \sum_{i=0}^{\infty} \frac{[(1-p')M]^i}{i!} e^{-(1-p')M} \cdot e^{-p'M} \\ &= e^{-p'M}. \end{aligned}$$

We average the probabilities over the number of interfering nodes in a region because of two reasons. First, it makes the problem much more tractable. Second, in our simulation experiments, we fix the number of competing nodes in a region (which is N) and then vary the location of the nodes to approximate the Poisson distribution, which is configurationally closer to our analytical model; the alternative would be to generate 2, 3, 4, ... nodes within one region, get the throughput for the individual configuration and then calculate the average, which is not practical.

Next we need to calculate the transition probability P_{il} from *idle* to *long*. If there are i nodes around node x, for such a transition to happen, one and only one node should be able to complete one successful four-way handshake while other nodes do not transmit. Let p_s denote the probability that a node begins a successful four-way handshake at each slot, we can then calculate P_{il} as follows:

$$P_{il} = \sum_{i=1}^{\infty} i p_s (1-p')^{i-1} \frac{M^i}{i!} e^{-M} = \sum_{i=1}^{\infty} p_s (1-p')^{i-1} \frac{M^{i-1}}{(i-1)!} M e^{-M} = p_s M e^{-p'M}.$$

To obtain the above result, we use the fact that the distribution of the number of nodes within R' does not depend on the existence of node x, because of the memoryless property of Poisson distribution. Up to this point, p_s is still an unknown quantity that we derive subsequently.

The transition probability from *idle* to *short1* is the probability that more than one node transmit RTS packets in the same slot; therefore, P_{is1} can be calculated as follows:

$$P_{is1} = \sum_{i=2}^{\infty} [1-(1-p')^i - ip'(1-p')^{i-1}] \frac{M^i}{i!} e^{-M} = 1-(1+Mp')e^{-p'M}.$$

Having calculated P_{ii}, P_{il} and P_{is1}, we can calculate P_{is2}, the transition probability from *idle* to *short2*, which is equal to $1 - P_{ii} - P_{il} - P_{is1}$. Let π_i, π_l, π_{s1} and π_{s2} denote the steady-state probabilities of states *idle*, *long*, *short1* and *short2* respectively. From Figure 1, we have

$$\begin{aligned} \pi_i P_{ii} + \pi_l + \pi_{s1} + \pi_{s2} &= \pi_i \\ \pi_i P_{ii} + 1 - \pi_i &= \pi_i \\ \pi_i &= \frac{1}{2-P_{ii}} = \frac{1}{2-e^{-p'M}}. \end{aligned}$$

The limiting probability Π_I, i.e., the long run probability that the channel around node x is found idle, can be obtained by:

$$\begin{aligned} \Pi_I &= \frac{\pi_i T_{idle}}{\pi_i T_{idle} + \pi_l T_{long} + \pi_{s1} T_{short1} + \pi_{s2} T_{short2}} \\ &= \frac{T_{idle}}{T_{idle} + P_{il} T_{long} + P_{is1} T_{short1} + P_{is2} T_{short2}}. \\ & (\pi_i P_{il} = \pi_l, \pi_i P_{is1} = \pi_{s1} \text{ and } \pi_i P_{is2} = \pi_{s2}) \end{aligned}$$

The relationship between p' and p is then:

$$p' = \frac{p T_{idle}}{T_{idle} + P_{il} T_{long} + P_{is1} T_{short1} + P_{is2} T_{short2}}. \quad (1)$$

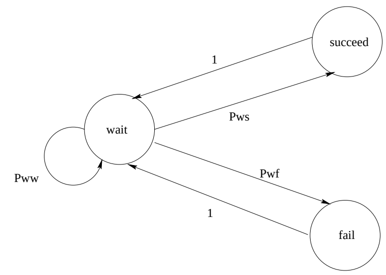


Figure 2. Markov chain model for a node

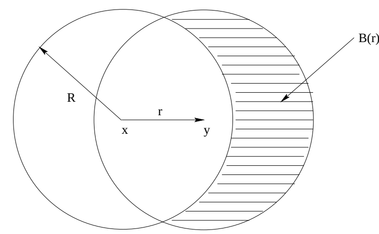


Figure 3. Illustration of "hidden" area

In the above equation, the probability that a node x starts successfully a four-way handshake in a time slot, p_s, is yet to be determined.

The states of a node x can be modeled by a three-state Markov chain, which is shown in Figure 2.

In Figure 2, *wait* is the state when the node defers for other nodes or backs off, *succeed* is the state when the node can complete a successful four-way handshake with other nodes, and *fail* is the state when the node initiates an unsuccessful handshake. For simplicity, we regard *succeed* and *fail* as the states when two different kinds of *virtual* packets are transmitted and their lengths are:

$$\begin{aligned} T_{succeed} &= T_{long} = l_{rts} + l_{cts} + l_{data} + l_{ack} + 4\tau \\ T_{fail} &= T_{short2} = l_{rts} + l_{cts} + 2\tau. \end{aligned}$$

Obviously, the duration of a node in *wait* state T_{wait} is τ.

Because by assumption collision avoidance is enforced at each node, no node is allowed to transmit data packets continuously; therefore, the transition probabilities from *succeed* to *wait* and from *fail* to *wait* are both one.

To derive the transition probability P_{ws} from *wait* to *succeed*, we need to calculate the probability $P_{ws}(r)$ that node x successfully initiates a four-way handshake with node y at a given time slot when they are at a distance r apart. Before calculating $P_{ws}(r)$, we define $B(r)$ to be the area that is in the hearing region of node y but outside the hearing region of node x, i.e., the interfering region "hidden" from node x as the shaded area shown in Figure 3. $B(r)$ has been shown [7] to be:

$$B(r) = \pi R^2 - 2R^2 q\left(\frac{r}{2R}\right)$$

where $q(t) = \arccos(t) - t\sqrt{1-t^2}$.

Then $P_{ws}(r)$ can be calculated as:

$$P_{ws}(r) = P_1 \cdot P_2 \cdot P_3 \cdot P_4(r)$$

where

$$\begin{aligned} P_1 &= \text{Prob.}\{x \text{ transmits in a slot}\}, \\ P_2 &= \text{Prob.}\{y \text{ does not transmit in the time slot}\}, \\ P_3 &= \text{Prob.}\{\text{none of the terminals within } R \text{ of } x \\ &\quad \text{transmits in the same slot}, \\ P_4(r) &= \text{Prob.}\{\text{none of the terminals in } B(r) \\ &\quad \text{transmits for } (2l_{rts}+1) \text{ slots} \mid r\}. \end{aligned}$$

The reason for the last term is that the vulnerable period for an RTS is only $2l_{rts}+1$, and once the RTS is received successfully by the receiving node (which can then start sending the CTS), the probability of further collisions is assumed to be negligibly small.

Obviously, $P_1 = p'$ and $P_2 = (1-p')$. On the other hand, P_3 can be obtained by

$$\begin{aligned} P_3 &= \sum_{i=0}^{\infty}(1-p')^i \frac{(\lambda\pi R^2)^i}{i!} e^{-\lambda\pi R^2} \\ &= \sum_{i=0}^{\infty}(1-p')^i \frac{N^i}{i!} e^{-N} = e^{-p'N}. \end{aligned}$$

Similarly, the probability that none of the terminals in $B(r)$ transmits in a time slot is given by

$$\begin{aligned} p_4(r) &= \sum_{i=0}^{\infty}(1-p')^i \frac{(\lambda B(r))^i}{i!} e^{-\lambda B(r)} \\ &= e^{-p'\lambda B(r)}. \end{aligned}$$

Hence, $P_4(r)$ can be expressed as

$$P_4(r) = (p_4(r))^{2l_{rts}+1} = e^{-p'\lambda B(r)(2l_{rts}+1)}.$$

Given that each sending node chooses any one of its neighbors with equal probability and that the average number of nodes within a region of radius r is proportional to r^2, the probability density function of the distance r between node x and y is

$$f(r) = 2r, \quad 0 < r < 1.$$

where we have normalized r with regard to R by setting $R = 1$.

Now we can calculate P_{ws} as follows:

$$\begin{aligned} P_{ws} &= \int_0^1 2rP_{ws}(r)dr \\ &= 2p'(1-p')e^{-p'N}\int_0^1 re^{-p'\lambda B(r)(2l_{rts}+1)}dr \\ &= 2p'(1-p')e^{-p'N} \cdot \\ &\quad \int_0^1 re^{-p'N[1-2q(r/2)/\pi](2l_{rts}+1)}dr. \end{aligned}$$

From the Markov chain shown in Figure 2, the transition probability P_{ww} that node x continues to stay in *wait* state in a slot is just $(1-p')e^{-p'N}$, i.e., it does not initiate any transmission and there is no node around it initiating a transmission. Let π_s, π_w and π_f denote the steady-state probability of state *succeed*, *wait* and *fail*, respectively. From Figure 2, we have

$$\begin{aligned} \pi_w P_{ww} + \pi_s + \pi_f &= \pi_w \\ \pi_w P_{ww} + 1 - \pi_w &= \pi_w \\ \pi_w &= \frac{1}{2 - P_{ww}} = \frac{1}{2-(1-p')e^{-p'N}}. \end{aligned}$$

Therefore, the steady-state probability of state *succeed* π_s can be calculated as:

$$\pi_s = \pi_w P_{ws} = \frac{P_{ws}}{2-(1-p')e^{-p'N}}. \quad (2)$$

We should note that π_s is just the previous unknown quantity p_s in Equation (1). Combining Equations (1) and (2) together, we get a complex relationship between p and p'. However, given p, p' can be computed easily with numerical methods.

Accordingly, the throughput Th is:

$$\begin{aligned} Th &= \frac{\pi_s \cdot l_{data}}{\pi_w T_w + \pi_s T_s + \pi_f T_f} \\ &= l_{data}\pi_s[\tau\pi_w + (l_{rts}+l_{cts}+2\tau)(1-\pi_w-\pi_s) \\ &\quad +(l_{rts}+l_{cts}+l_{data}+l_{ack}+4\tau)\pi_s]^{-1}. \quad (3) \end{aligned}$$

To apply our analysis to MAC protocols in which perfect collision avoidance is not enforced, e.g., the IEEE 802.11 MAC protocol, we propose a simple though not rigorous extension of the analysis. We can add another state to the Markov chain for the node model (ref. Figure 2) whose duration is $l_{rts} + l_{cts} + l_{data} + 3\tau$. This is a *pseudo-succeed* state in which an RTS-CTS-data handshake takes place without acknowledgment due to collisions, i.e., it is a state derived from the *succeed* state of the perfect collision avoidance protocol. We use an "imperfectness factor" β to model the deviatory behavior of the protocol, given that different MAC protocols may have different values of β. The transition probability from *wait* to the *pseudo-succeed* state is then βP_{ws}, and the transition probability from *wait* to *succeed* is $(1-\beta)P_{ws}$. Hence, the modified formula for throughput is simply:

$$\begin{aligned} Th &= (1-\beta)l_{data}\pi_s[\tau\pi_w + (l_{rts}+l_{cts}+l_{data}+ \\ &\quad l_{ack}+4\tau)(1-\beta)\pi_s + (l_{rts}+l_{cts}+2\tau)(1-\pi_w \\ &\quad -\pi_s) + (l_{rts}+l_{cts}+l_{data}+3\tau)\beta\pi_s]^{-1}. \quad (4) \end{aligned}$$

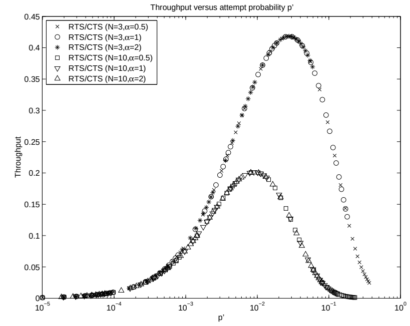
When the deviatory factor β equals zero, Equation (4) is reduced to Equation (3).

3 Numerical Results

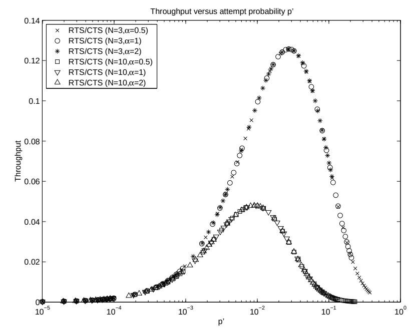
The performance of non-persistent CSMA protocol in multi-hop networks has been analyzed by Wu and Varshney [8]. In their model, the authors assume that there is a separate channel over which acknowledgments are sent in zero time and without collisions. In practice, the performance of CSMA with acknowledgments would be much worse than the predictions of the idealized model, as predicted by Tobagi and Kleinrock [9].

We present results when either relatively large data packets or relatively small data packets are sent. Let τ denote the duration of one time slot. RTS, CTS and ACK packets last 5τ. As to the size of data packets, we consider two cases. One case corresponds to a data packet that is much larger than the aggregate size of RTS, CTS and ACK packets. The other case corresponds to a data packet being only slightly larger than the aggregate size of RTS, CTS and ACK packets. In the latter case, which models networks in which radios have long turn-around times and data packets are short, it is doubtful whether a collision avoidance scheme should be employed at all, because it represents excessive overhead. We first calculate throughput with different values of α, which we define as the ratio between the circular region including nodes affected by an RTS/CTS handshake and the largest possible circular region in which nodes are guaranteed to be connected with one another. We find that, though the relationship between the ready probability p and transmission attempt probability p' under different values of α might be somewhat different, the throughput is largely unaffected by α, which is shown in Figure 4. In Figure 4, N is the average number of nodes that compete against one another to access the shared channel. Thus, the burden of estimating α is relieved in our model, and we can focus on the case in which $\alpha = 1$ thereafter. However, as a side effect of not knowing the actual α that should be used, the relationship between p' and throughput may not agree with the simulations. However, for our purposes this is not a problem, because we are interested in the saturated throughput only.

Figure 5 compares the throughput of collision avoidance against that of CSMA with different values of N and data packet lengths and we can make the following observations from the above results. When data packet is long, the throughput of CSMA is very low, even for the case in which only $N = 3$ nodes are competing for the shared channel. In comparison, the RTS/CTS scheme can achieve much higher throughput even when the average number of competing nodes is 10. When a data packet is not very long and the overhead of the collision avoidance and handshake seems to be rather high, collision avoidance can still achieve marginally better throughput than CSMA. We need to emphasize that the performance of the actual CSMA will



(a) long data packet: $l_{data} = 100\tau$



(b) short data packet: $l_{data} = 20\tau$

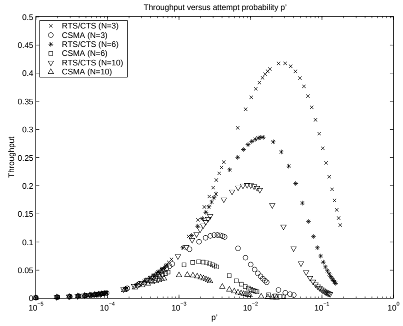
Figure 4. α's influence ($l_{rts} = l_{cts} = l_{ack} = 5\tau$)

be much worse than the idealized model we have used for comparison purposes, because of the effect of acknowledgments.

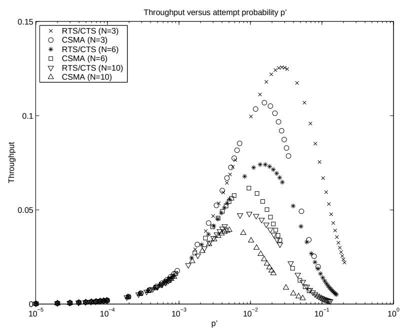
Despite the advantage of collision avoidance, the throughput still degrades rapidly with the increase of N. This is also evident of low values of p' as shown in Figure 5. This is due the fact that nodes are spending much more time on collision avoidance and backoff. When N increases, p' decreases much slower to achieve optimum throughput, which already decreases. This shows that collision avoidance becomes more and more ineffective when the number of competing nodes within a region increases, even though these nodes are quite "polite" in their access to the shared channel. This is also different from a fully-connected network, in which the maximum throughput is largely indifferent to the number of nodes within a region [3].

4 Simulation Results

We use GloMoSim 2.0 [10] as the network simulator. Direct sequence spread spectrum (DSSS) parameters are used throughout the simulations. The raw channel bit rate



(a) long data packet: $l_{data} = 100\tau$



(b) short data packet: $l_{data} = 20\tau$

Figure 5. Throughput comparison ($l_{rts} = l_{cts} = l_{ack} = 5\tau$)

is 2Mbps. We use a uniform distribution to approximate the Poisson distribution used in our analytical model, because the latter is mainly used to facilitate our derivation of analytical results. To be specific, we place nodes in concentric circles or rings. That is, given that a node's transmitting and receiving range is R and that there are on average N nodes within this circular region, we place N nodes in a circle of radius R, subject to a uniform distribution. Because there are on average 2^2N nodes within a circle of radius $2R$, we place $2^2N - N = 3N$ nodes outside the previous circle of radius R but inside the concentric circle of radius $2R$, i.e., the ring with radii R and $2R$, subject to the same uniform distribution. Then $3^2N - 2^2N = 5N$ nodes can be placed in an outer ring with radii $2R$ and $3R$. Given that we cannot generate an infinite network model, we just focus our attention on the performance of the innermost N nodes. According to our experiments, we find that nodes that are outside the concentric circles of radius $3R$ almost have no influence on the throughput of the innermost N nodes, i.e., boundary effects can be safely ignored when the circular network's

	τ	l_{rts}	l_{cts}, l_{ack}	l_{data}
actual time	21μsec	272μsec	248μsec	6032μsec
normalized	1	13	12	287

Table 1. Equivalent configuration parameters for analytical model

radius is $3R$. Accordingly, we present only the results for a circular network of radius $3R$.

The backoff timer in the IEEE 802.11 MAC protocol is drawn from a uniform distribution whose upper bound varies according to the estimated contention level, i.e., a modified binary exponential backoff. Thus, p' takes on dynamic values rather than what we have assumed in the analytical model. Accordingly, we expect that the IEEE 802.11 MAC protocol will operate in a region, while our analysis gives only average performance. In addition, the IEEE 802.11 MAC protocol cannot ensure collision-free transmission of data packets, even under the assumption of perfect carrier sensing and collision avoidance, which has been discussed in [2].

In our simulation, each node has a constant-bit-rate (CBR) traffic generator with data packet size of 1460 bytes, and one of its neighbors is randomly chosen as the destination for each packet generated. All nodes are always backloged. For comparison purposes, we map these simulational parameters to equivalent parameters in our analytical model and they are shown in Table 1.

We run both analytical and simulation programs with N = 3, 5 and 8. Though we have not tried to characterize how the performance of the IEEE 802.11 MAC protocol is distributed in the region of values taken by p', we do have generated 50 random topologies that satisfy the uniform distribution and then get an average transmission probability and throughput for the N nodes in the innermost circle of radius R for each configuration. The results are shown in Figure 6, in which the centers of rectangles are the mean values of p' and throughput and their half widths and half heights are the variance of p' and throughput respectively. These rectangles roughly describe the operating regions of IEEE 802.11 MAC protocol with the configurations we are using.

Figure 6 clearly shows that IEEE 802.11 cannot achieve the performance predicted in the analysis of correct collision avoidance, but may well outperform the analysis with the same p' for some configurations, especially when N is small. On first thought, it may seem contrary to intuition, given that IEEE 802.11 cannot ensure collision-free data packet transmissions and should always perform worse than analysis results. In fact, the exceedingly high throughput is largely due to the unfairness of the binary exponential backoff (BEB) used in IEEE 802.11. In BEB, a node that just succeeds in sending a data packet resets its contention window to the minimum value, through which it may gain access to the channel again much earlier than other surrounding nodes. Thus, a node may monopolize the channel for a very long time during which there is no contention loss and throughput can be very high for a particular node, while other nodes suffer starvation. We also find that when N increases, the variance of p' and throughput becomes smaller. Thus, the fairness problem is less severe when there are more nodes competing in a shared channel.

Due to the inherent deficiency of the BEB scheme used in the IEEE 802.11 MAC protocol, we investigate a simple variant of this protocol in which the contention window (*CW*) is fixed and then the backoff timer is generated from a uniform distribution with values ranging between 0 and *CW*. We vary the *CW* to get different values of p' and throughput. Though this modified protocol is not fine tuned to the actual number of neighbors that a node may have and thus is not able to deliver the best performance, it is still a much fairer scheme that reflects more realistically how well a contention-based MAC protocol may perform. In order to have a fair comparison of this scheme with the original IEEE 802.11, we reuse the aforementioned node configurations. The *CW* used in our simulations are tabulated in Table 2 and the simulation results are shown in Figure 7. For clarity, Figure 7 shows only the operating regions (shown in rectangles) of the modified IEEE 802.11, without showing details of how each set of the 50 configurations performs. In addition, the median values of p' and throughput are drawn in Figure 7 to show how the throughput is affected by p' or *CW*, where a larger *CW* means a smaller p'.

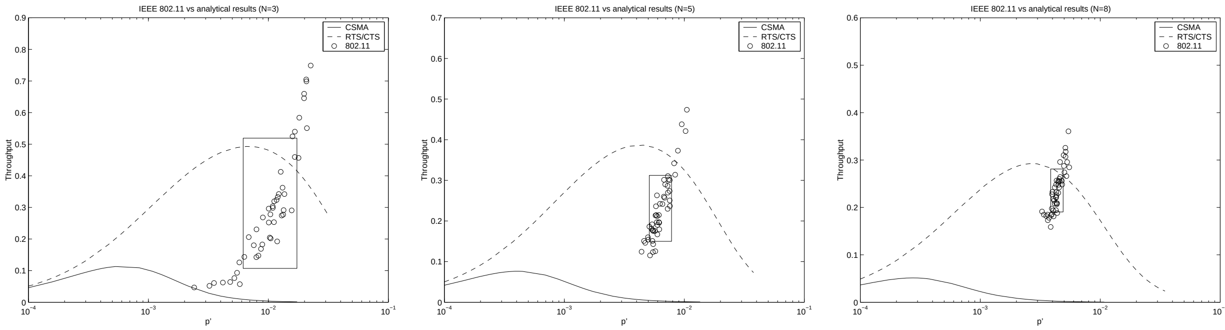
Comparing Figures 6 and 7, it is very clear that the modified IEEE 802.11 protocol with fixed *CW* has a smaller variance of throughput than that in the original protocol and thus is much fairer. We can also see the degraded performance in the fixed *CW* scheme due to more contention, especially when N is small. Given that these two protocols cannot ensure that data packets are transmitted free of collisions, the throughput can deviate substantially from what is predicted in the analysis. To demonstrate this, we also collect statistics about the number of transmitted RTS packets that will lead to ACK timeout due to collision of data packets as well as the total number of transmitted RTS packets that can lead to either an incomplete RTS-CTS-data handshake or a successful four-way handshake. Then we calculate the ratio of these two numbers and tabulate the results in Table 3. This table clearly shows that much of the precious channel resource is wasted in sending data packets that cannot be successfully delivered. In addition, in order to decrease the percentage of channel resource wasted due to collisions, a larger contention window should be chosen to artificially decrease the transmission probability of nodes which at the same time lead to longer time wasted in waiting. This is a very typical behavior of collision-avoidance protocols, especially those protocols that do not have a correct collision avoidance scheme. The possibility of collisions of data packets with other packets places a limit on the maximum achievable throughput, which can be significantly lower than the theoretical results that assume a perfect collision avoidance.

A close observation of Figures 6 and 7 also reveals that, the gap in maximum throughput between analytical and simulation results decreases when N increases. This can be explained as follows. When the number of direct competing nodes N increases, the number of indirect competing nodes (hidden terminals, $3N$ on average) also increases, which makes nodes implementing a perfect collision avoidance protocol spend much more time in deferring and backing off to co-ordinate with both direct and indirect competing nodes to avoid collisions. Therefore, much of the gain of perfect collision avoidance is lost and possible spatial reuse is also reduced in congested area, which makes a perfect collision avoidance protocol work only marginally better than an imperfect one. This observation could not be predicted from previous analytical models or simulations focusing on fully-connected networks or networks with only a limited number of hidden terminals [3–5].

The percentage shown in Table 3 is in fact the β in our extended analysis for imperfect collision avoidance protocols. Using these values, we compare the performance of the IEEE 802.11 protocol with that of the adjusted analysis obtained from Equation (4), which is shown in Figure 8. In Figure 8, we only show the results for small values of N as it is not quite meaningful to do the adjustment for large values of N due the reason stated above. Figure 8 shows that the extended analysis is a rather good approximation of the actual performance of the IEEE 802.11 protocol though the latter has larger variation in throughput (possibly due to its inherent fairness problems).

5 Conclusion

We have used a simple model to derive the saturation throughput of MAC protocols based on an RTS-CTS-data-ACK handshake in multi-hop networks. The results show that these protocols outperform CSMA protocols, even when the overhead of RTS/CTS exchange is rather high, thus showing the importance of correct collision avoidance in random access protocols. It is also shown that the overall performance of the sender-initiated collision avoidance scheme degrades rather rapidly when the number of competing nodes allowed within a region increases, in contrast to the case of fully-connected networks and networks with limited hidden terminals reported in the literature [3–5], where throughput remains almost the same for a large num-

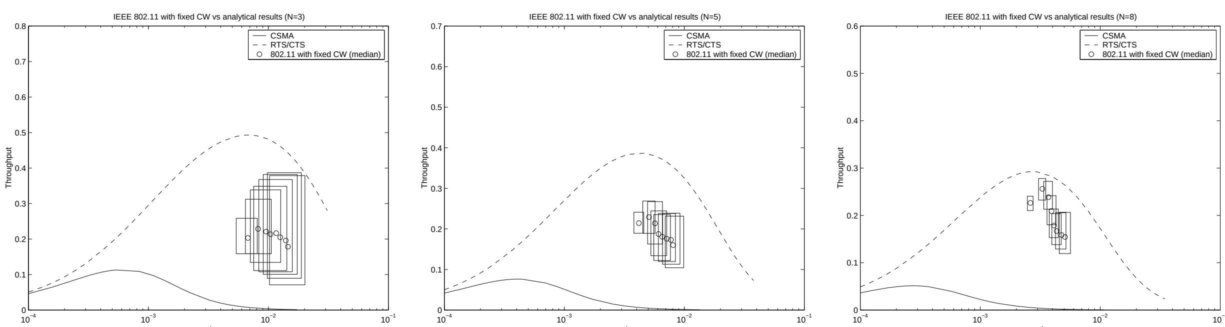


(a) $N = 3$ (b) $N = 5$ (c) $N = 8$

Figure 6. Performance comparison of IEEE 802.11 with analytical results

	CW1	CW2	CW3	CW4	CW5	CW6	CW7	CW8
N=3	20	40	80	160	320	640	1280	2560
N=5	30	60	120	240	480	960	1920	3840
N=8	50	100	200	400	800	1600	3200	6400

Table 2. Contention window (CW) used in simulations

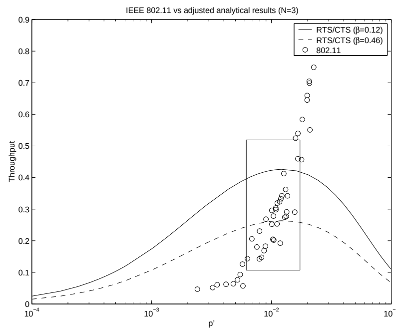


(a) $N = 3$ (b) $N = 5$ (c) $N = 8$

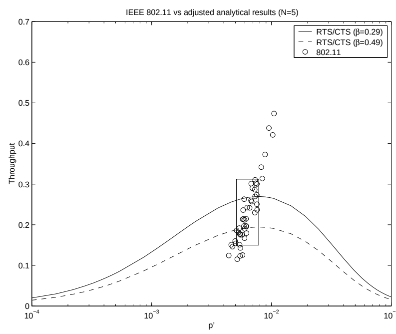
Figure 7. Performance comparison of IEEE 802.11 (fixed CW) with analytical results

	original	CW1	CW2	CW3	CW4	CW5	CW6	CW7	CW8
N=3, mean	0.29	0.49	0.46	0.43	0.41	0.39	0.33	0.23	0.14
N=3, std	0.17	0.20	0.18	0.17	0.17	0.17	0.15	0.11	0.07
N=5, mean	0.39	0.56	0.54	0.53	0.51	0.47	0.37	0.25	0.15
N=5, std	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.07	0.04
N=8, mean	0.44	0.59	0.58	0.56	0.53	0.44	0.33	0.21	0.14
N=8, std	0.06	0.07	0.07	0.06	0.06	0.06	0.05	0.04	0.07

Table 3. Percentage of ACK timeout in BEB scheme and fixed CW



(a) $N = 3$



(b) $N = 5$

Figure 8. Performance comparison of IEEE 802.11 with adjusted analytical results

ber of nodes. Simulation experiments with the IEEE 802.11 MAC protocol validate these observations and show that the IEEE 802.11 MAC protocol can suffer severe degradation in throughput due to its inability to avoid collisions between data packets and other packets.

Based on both analytical and simulation results, we observe that there are some possible ways to improve the throughput of a sender-initiated collision-avoidance protocol in ad hoc networks. First of all, the simulation results show that it is very important to ensure correct collision avoidance. Using the longer CTS packets proposed in [2], the IEEE 802.11 MAC protocol can lead to much better performance in throughput.

Another obvious way to improve the performance of the IEEE 802.11 protocol is to reduce τ, which includes carrier sensing delay and transmit to receive turnaround time, so as to enlarge the ratio of data packet transmission time to τ. In effect, this implies reducing the transmission power of the nodes and reducing the length of control overhead. Given that RTS and CTS packets cannot be reduced in length and arguably the CTS needs to be lengthened to be sent as a busy tone, the latter requires using piggyback acknowledgments or making acknowledgments optional.

Because the optimum value of p' changes with the number of competing nodes within a region, it is necessary to have an adaptive algorithm to achieve optimum performance when the number of active nodes within a region changes. Given that the original BEB scheme has inherent fairness problem and the fixed contention window does not adapt well, it is fair to say that there is still much work left to be done on this topic.

References

[1] IEEE Computer Society LAN MAN Standards Committee, ed., *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE Std 802.11-1997, The Institute of Electrical and Electronics Engineers, New York, 1997.

[2] C. L. Fullmer and J. J. Garcia-Luna-Aceves, “Solutions to Hidden Terminal Problems in Wireless Networks,” in *Proc. of ACM SIGCOMM ’97*, 1997.

[3] G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed Coordination Function,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, Mar. 2000.

[4] F. Cali, M. Conti, and E. Gregori, “IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1774–1786, Sept. 2000.

[5] F. Cali, M. Conti, and E. Gregori, “Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit,” *IEEE/ACM Transactions on Networking*, vol. 8, pp. 785–799, Dec. 2000.

[6] H. S. Chhaya and S. Gupta, “Throughput and Fairness Properties of Asynchronous Data Transfer Methods in the IEEE 802.11 MAC Protocol,” in *Proc. of PIMRC ’95*, 1995.

[7] H. Takagi and L. Kleinrock, “Optimal Transmission Range for Randomly Distributed Packet Radio Terminals,” *IEEE Transactions on Communications*, vol. 32, no. 3, pp. 246–57, 1984.

[8] L. Wu and P. Varshney, “Performance Analysis of CSMA and BTMA Protocols in Multihop Networks (I). Single Channel Case,” *Information Sciences, Elsevier Sciences Inc.*, vol. 120, pp. 159–77, 1999.

[9] F. A. Tobagi and L. Kleinrock, “The Effect of Acknowledgment Traffic on the Capacity of Packet-Switched Radio Channels,” *IEEE Trans. on Communications*, vol. 26, no. 6, pp. 815–826, 1978.

[10] X. Zeng, R. Bagrodia, and M. Gerla, “GloMoSim: a Library for Parallel Simulation of Large-scale Wireless Networks,” in *Proc. of the 12th Workshop on Parallel and Distributed Simulations*, May 1998.