An analytical model of delay in multi-hop wireless ad hoc networks

**E. Ghadimi · A. Khonsari · A. Diyanat ·
M. Farmani · N. Yazdani**

Abstract Several analytical models of different wireless networking schemes such as wireless LANs and meshes have been reported in the literature. To the best of our knowledge, all these models fail to address the accurate end-to-end delay analysis of multi-hop wireless networks under unsaturated traffic condition considering the hidden and exposed terminal situation. In an effort to gain deep understanding of delay, this paper firstly proposes a new analytical model to predict accurate media access delay by obtaining its distribution function in a single wireless node. The interesting point of having the media access delay distribution is its generality that not only enables us to derive the average delay which has been reported in almost most of the previous studies as a special case but also facilitates obtaining higher moments of delay such as variance and skewness to capture the QoS parameters such as jitters in recently popular multimedia applications. Secondly, using the obtained single node media access delay distribution, we extend our modeling approach to investigate the delay in multi-hop networks. Moreover, probabilities of collisions in both hidden and exposed terminal conditions have been calculated. The validity of the model is demonstrated by comparing results predicted by the analytical model against those obtained through simulation experiments.

E. Ghadimi
School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden

A. Khonsari · M. Farmani · N. Yazdani
Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

A. Khonsari
School of Computer Science, IPM, Tehran, Iran

A. Diyanat (✉)
Department of Electrical Engineering, Sharif University, Tehran, Iran
e-mail: Abolfazl.diyanat@gmail.com

1 Introduction

Wireless ad hoc networks (WANs) consist of several numbers of nodes which communicate with one another in an unattended collaborative manner. Any two nodes within the given Euclidean distance can communicate directly. In such networks, a data packet may be forwarded by several intermediate nodes till it reaches the destination. The wireless media in these networks is a shared and scarce resource therefore, a well-defined MAC protocol plays a prominent role in performance of a WAN. Due to lack of a centralized control, MAC protocols must be implemented in a distributed manner in ad hoc networks. Distributed coordination function (DCF) access mechanism of IEEE 802.11 as the basic standard for MAC layer has gained common popularity in ad hoc networks Takano and Liou [1]. Although the basic characteristics of IEEE 802.11 DCF are well understood, this paper suggests required analytical models to evaluate dynamic performance of WANs in terms of multi-hop delay analysis. This aim is achieved by proposing a local queueing system to address delay analysis of multi-hop wireless networks under unsaturated traffic condition.

The overall end-to-end packet delay consists of two types of latencies: medium access delay and queueing delay. Medium access delay comprises of the latencies for

data transmissions and retransmissions. On the other hand, queueing delay is the amount of time that the data packet waits until receives service from the MAC layer interface. To derive multi-hop delay, we employ a Markov chain model to analyze the probability of transmission at each node in an arbitrary slot and derive the wireless channel access delay. Like Ozdemir and McDonald [2], Zhai et al. (2004), and Shabdiz and Subramanian [3], the arrival process at each node is assumed to be an independent and identically distributed (*iid*) Poisson process. Moreover, the distribution of service times are modeled as General Independent processes. Hence, each node is taken to be an M/G/1 queue for which we derive the queuing delay. Thus, the average single hop delay namely, the average delay from the time a packet arrives at some node to the time it is successfully received by the next hop node can be calculated. For this reason, we predict the accurate media access delay by obtaining its distribution function in a single wireless node. Calculating the media access delay distribution enables us to derive the moments of delay (e.g., first moment and variance). These parameters consequently lead us to the average waiting time in M/G/1 queueing system. Other useful parameters such as QoS jitters with widespread use in recently popular multimedia applications can be obtained by higher moments of the service time.

IEEE 802.11 DCF mechanism suffers from problems concerning hidden and exposed nodes Marsic [4]. While use of RTS/CTS message passing solves the problem of hidden nodes, the exposed terminal problem is left uncovered in current implementations of the IEEE 802.11 standard.

For the first time, in this paper we discuss the performance impact of exposed terminals in IEEE 802.11 ad hoc networks under finite load conditions. In this context, we derive accurate analytical models for the media access delay for IEEE 802.11 ad hoc networks in finite load conditions with and without exposed terminals. Thus, the model extends from analyzing the single-hop average packet delay to evaluating the end-to-end packet delay in multi-hop ad hoc networks. The major contribution of this paper is to analyze the average packet delay for IEEE 802.11 DCF by using an analytic model under finite load traffic in multi-hop ad hoc networks for which no accurate model has been reported in the literature.

The remainder of this paper is organized as follows. Section 2 discusses the related work in this avenue. Section 3 describes the system model which encompasses models to derive the Markov chain state transition probabilities, packet service time and queueing time in single-hop ad hoc networks. Section 4 includes the calculation of probabilities of collisions in both hidden and exposed terminal conditions. Moreover, in this section, we propose our modeling approach to investigate the delay in multi-hop

networks. Section 5 applies simulation and statistical analysis to validate the analytical results. Finally, Section 6 presents concluding remarks and provides guidelines for future work.

2 Related work

Analytical modeling and performance evaluation of the MAC layer are fascinating issues to pursue in the literature. In Li [5], authors have studied the relation between capacity of each node and characteristics of MAC layer IEEE 802.11 via simulation experiments. They have considered different traffic patterns for various network topologies such as single cell, chain and random network. An estimate of the expressions for one-hop capacity and upper bound of per-node throughput is obtained using the simulation results.

In Bianchi [6] and Wu [7], performance of the IEEE 802.11 was evaluated in simplified scenarios. These scenarios were performed under saturation condition, i.e. each node in the network always has packet to transmit. But saturation assumption for the network seems inappropriate, especially when we know that the number of packets to be transmitted depends on the incoming traffic pattern.

Several other models have been proposed to investigate the performance of IEEE 802.11 under unsaturated traffic conditions Bianchi [6], Ergen and Varajya [8]. Although the discrete-time Markov chain model introduced in Ergen and Varajya [8] takes the unsaturated traffic condition into account, it does not adopt any queuing system to model the MAC buffer (i.e. the MAC buffer length is assumed zero).

Analytical modeling of IEEE 802.11 DCF using queuing system has been studied in Tickoo and Sikdar [9], Ozdemir and McDonald [2], and Zhai (2004). The proposed model in Tickoo and Sikdar [9] is based on a G/G/1 queue. From the computational complexity point of view, this depends on numerous approximated parameters, such as the probability that each node has no packet to transmit. Thus in this model, the results obtained from simulation show much variation than that from the analysis.

Performance analysis of the MAC layer in the finite queue size condition has been deliberated in Ozdemir and McDonald [2] and Zhai (2004), both of which are based on the M/G/1/K queuing system. The two models differ in how the service time distribution is derived. While Ozdemir et al. use the Markov-modulated general independent model in Ozdemir and McDonald [2], authors of Zhai (2004) calculate the service time distribution using the transform function.

In Bisnik and Abouzeid [10], Bisnik et al. studied the performance of wireless ad hoc networks in terms of throughput and packet delay in single and multi hop

scenarios. They proposed an analytical model based on G/G/1 queuing networks and used the diffusion approximation in order to evaluate closed form expressions for the average end-to-end delay. However, similar to Tickoo and Sikdar [9], their work depends on approximated parameters. Likewise, Bisnik et al. Bisnik and Abouzeid (2006) characterized the average delay and capacity in random access MAC based wireless mesh networks (WMNs) and modeled residential area WMNs as an open G/G/1 queuing network using the diffusion approximation method to indicate the scalability of WMNs performance with the number of mesh routers and clients.

In Tickoo and Sikdar [11], Tickoo et al. devised an analytic model to evaluate the queuing delays for the WANS using IEEE 802.11 DCF model. They have used G/G/1 queuing model for each node and have extended their study under consideration of 802.11e standard wherein a number of packets may be transmitted in a burst once the channel is accessed. Although Tickoo and Sikdar [11] considers unsaturated traffic model. However, do not address an accurate model for the multi-hop transmissions. Also, authors do not take into account hidden and exposed nodes problems as the main drawback of IEEE 802.11 DCF mechanism.

In Vassis and Kormentzas [12], Vassis et al. evaluated the performance effects of exposed terminals in IEEE 802.11 ad hoc networks under finite load conditions. This paper used the model in Tobagi and Kleinrock [13] to derive total delay and network utilization parameters. In spite of considering the exposed terminal problem, only single-hop paths have been investigated in this paper. Moreover, it has been assumed that every node in the network is visible to all the other nodes; however, this constraint may not be valid in practice.

In a way different from other methods, Zheng (2006) introduces an analytical model to evaluate the performance of DCF in imperfect wireless channels. Considering various incoming traffic loads, error conditions in wireless channels and its impact on the performance of the network is addressed through analysis and simulation experiments.

But the most related model to our analysis is that proposed by Shabdiz et al. in Shabdiz and Subramanian [3]. In this work, they proposed approximate analytical models for the throughput performance of single-hop and multi-hop ad hoc networks. Using the Markov chain introduced in Bianchi [6], the MAC layer behavior has been modeled under the DCF mode and the RTS/CTS mechanism. Closed-form expressions of packet delay and throughput in single hop network have also been presented in the paper. The supposed queuing system is M/G/1 and in order to estimate the queuing delay, the probability distribution function and first moment of the service time have been derived for the single hop scenario. For multi-hop network,

the authors took into account the hidden node problem and derived the required probabilities of collision occurrence. However, the total end-to-end delay in multi-hop networks has not been addressed and only the approximate throughput for the multi-hop condition has been calculated.

3 System model

A number of tools are available to measure the performance of a network; namely Analysis, Simulation, and Experiment. We typically start using analysis to estimate network performance using a mathematical model of the network. Analysis provides approximate performance numbers and gives insight into how different factors affect performance. Analysis also allows an entire family of networks with varying parameters and configuration to be evaluated at once by deriving a set of equations that predicts the performance of the entire family. However, analysis usually involves making a number of approximations that may affect the accuracy of results. There has been a lot of effort in the literature to make analytical models more precise in order to include as many parameters of the networks as possible. However, so far there has not been any analytical model to incorporate precisely all the parameters of the underlying system completely.

In this section, we describe the analytical model to calculate message latency using queuing theory. The model is based on a wireless network which consists of n stationary nodes sharing a common medium, and packets are transmitted from sources to destinations with an already known routing algorithm. The nodes use the IEEE 802.11 DCF (RTS/CTS mode) as the MAC protocol. Each node has an infinite buffer for storing packets. We assume that the aggregate offered traffic load in the wireless channel is generally distributed and therefore, packets arrive to a node with the rate of λ_g. The packet size transmitted in the wireless medium is also generally distributed with a mean value of P bits. The channel is assumed to be error-free. Roughly speaking, a packet transmission is considered to be successful if there are no other simultaneous packet transmissions.

The mean message latency is composed of the mean network latency, A_s, that is the transmission and retransmission time of the data across the network, and the mean queuing delay which is the waiting time seen by message at the interface queue, A_w. Therefore, we have:

Latency $= A_s + A_w$

A summary of the notation used in the derivation of the model is provided in the Appendix 1. We employ a Markov chain model to analyze the probability of transmission at each node in an arbitrary slot and derive the wireless

channel access delay. Then, we model each node as an M/G/1 queue and derive the average queuing delay at a node for both, single hop and multi hop scenarios. The modeling of the behavior of each node follows the 802.11 MAC protocol. Readers are referred to Marsic [4] for more details. The Markov chain model which describes transmission state of a node evolves from the model presented in Bianchi [6] as shown in Fig. 1.

In the system, each node alternates between busy and idle states. When the traffic is non-saturating, a node is not in backoff state all the time and additional states are necessary. During a busy period, the node executes the RTS/CTS protocol, transmits its data and receives an acknowledgment. If it has multiple frames to send, it may contend and transmit more than one during the same busy period. After each transmission, a node goes into backoff stage; other nodes continue counting down the backoff time according to the IEEE 802.11 standard (Takano and Liou [1]). The state of the channel can be characterized by two probabilities: $p_{tr}(n)$, defined as the probability that at least one of n nodes transmits a packet in a random slot, and $p_s(n)$, the probability that there was a successful transmission, given that at least one node transmitted a packet. Let τ be defined as the probability that a node transmits a packet in an arbitrarily chosen slot. According to Fig. 1, the IDLE

state represents the condition in which the node has no packet to transmit. If a packet arrives at the node with an empty queue (i.e., in the IDLE state), the node senses the medium. If the medium is idle the node transits to state FirstTR and sends the packet immediately. The probability of failure in the transmission given that a node sends a packet is denoted by c. Further, let q denote the probability that the node buffer becomes empty after each particular node finishes current packet processing (either transmit successfully or eventually drop it after maximum possible retransmission as defined in the IEEE 802.11 standard Takano and Liou [1]). Also, let $p_{nk}(t)$ be the probability of having no packet arrival at a given node during the period of t. In other words, $p_{nk}(t)$ is the probability of generating no packets in time duration t. As we mentioned before, the arrival process at each node is assumed to be an independent and identically distributed (iid) Poisson process with the rate of λ_g packets per second. Therefore, for a node in transmission mode, the probability of having no packet to send in time duration of t is equal by:

$$p_{nk}(t) = \frac{(\lambda_g t)^0}{0!} e^{-\lambda_g t} \tag{1}$$

Now assume a network of nodes with empty queues at the start of the transmission. Commencing from the IDLE

Fig. 1 State transition diagram of a single node using IEEE 802.11 standard
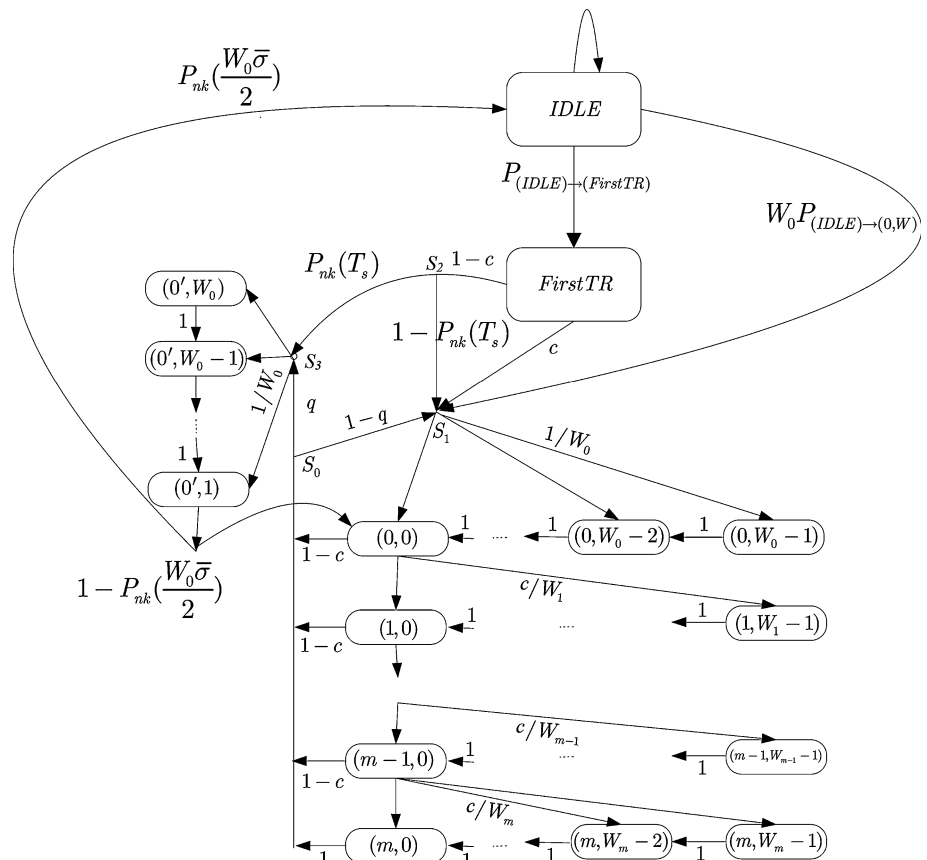
state, when a given node has a packet to transmit, it transits to FirstTR state. Considering the state FirstTR, if the medium is sensed idle, the node immediately sends the packet. If no collision is detected by the node (i.e., with probability $1 - c$), it will enter state S_2. Then, dependent upon having any packet in buffer (with probability $1 - p_{nk}$ (T_s)) or not, the node goes to either of states S_1 or S_3 (we will explain these two states later). On the other hand, in state FirstTR if the medium is sensed busy or the transmitted packet collides with others (with probability c), the node goes to check point S_1 and then enters the backoff state, where it is transmitted after the backoff timer reaches 0.

A backoff stage represents as the tuple (i, w), here i denotes the level of backoff stage, i.e. it is actually the number of transmission attempts ($i = 0, 1, ..., m', ..., m$). Also, $W_i = 2^i W_0$ denotes the window size of level i (W_0 is the minimum of the contention window). Moreover, w is the backoff counter in the range of $[0, W_i - 1]$. For the values of $m' < i \leq m$, W_i is kept at its maximum value. Here m' is the maximum size of the contention window. In other words, When the node reaches this threshold in packet transmission, it fixes the window size at the maximum value $W_i = 2^{m'} W_0$. After that, on each collision node retransmits the packet until drops it on stage $m + 1$. So, m denotes the maximum number of retransmissions for a given packet. $p_{i,w}$ symbolizes the probability of being in state (i, w). After a node finishes processing a packet (say state S_0), it resets i to zero. In this occasion, if the packet queue is not empty, the node goes to S_1 state directly, i.e. it sets the counter and enters the backoff state. Otherwise (when the buffer is empty), the node sets its backoff counter and after expiration of the counter, based upon the situation, it goes to IDLE state (when the packet buffer is still empty) or contends with other nodes to transmit a packet in backoff stages (when at least one packet arrives during the countdown). For the sake of clarity, let us denote the latter scenario as state S_3. This issue is captured with an extra set of backoff states which is depicted on the left side of Fig. 1.

Transitions from state to state occur at the end of channel slots. Three types of channel slots are defined, each with different time durations: idle, fail, or success. The duration of a channel slot is the period of time that the channel stays in one state: idle, fail, or success. Suppose T_s is the time needed to complete a successful transmission. So, T_s includes the time to send RTS, CTS, ACK and data packets, plus the inter-frame time. Also, let T_c denote the time for a failed transmission. Consider σ as the system time slot duration and A_p as the time needed to transmit the payload. Assuming that all stations use the same channel access mechanism, T_s and T_c are defined as follows, where the RTS/CTS access mechanism is employed:

$$T_s = \text{DIFS} + \text{RTS} + \text{SIFS} + \text{CTS}$$
$$\quad + \text{SIFS} + H + A_p + \text{SIFS} + \text{ACK} + \sigma$$
$$T_c = \text{DIFS} + \text{SIFS} + \text{RTS} + \text{C}TS$$

H is header size of the MAC layer payload. Since the described state diagram is an embedded Markov chain, so the future state of the node given the present state is independent of the past states. From the steady-state distribution of the Markov chain, the balance equations of state transition can be obtained as follows: Let p_{S0} denote the probability of being in state S_0. From this state, we can reach S_3 with probability q. Consequently, a transition takes place from state S_3 to either state $\{(0' \ w)$, $(0', w + 1), ..., (0', W_0)\}$ with probability $\frac{1}{W_0}$. The remarkable point here is that being in one of states, $W_0 - w$ would lead us to the $(0', w)$ state with probability 1. So the probability of reaching state $(0', w)$ which is denoted by $p_{0',w}$ is given as follows:

$$p_{0',w} = \frac{W_0 - w}{W_0} q p_{S_0} \tag{2}$$

Similarly, we can calculate Eqs. 3 and 4 as follows:

$$p_{i,0} = c^i p_{0,0} \tag{3}$$

$$p_{i,w} = \frac{W_i - w}{W_i} c^i p_{0,0} \tag{4}$$

In the described Markov state diagram (Fig. 1), FirstTR is accessible from IDLE state. In fact, when the medium is sensed idle, the node transits from IDLE state to FirstTR state. Also, $1 - p_{tr} (n - 1)$ is the probability that an idle channel slot occurs given that the node in question is idle. Remember that the entire scenario above occurs when the node has a packet to send (at least one packet must be generated during the idle system slot σ, i.e. with probability $1 - p_{nk} (\sigma)$). So the transition probability of FirstTR state is given as below:

$$p_{\text{FirstTR}} = p_{IDLE}(1 - p_{tr}(n - 1))(1 - p_{nk}(\sigma)) \tag{5}$$

Consider a scenario in which the node passes the extra set of backoff states (which is depicted in the left side of Fig. 1) and after that, it still has no packet in its buffer. In such occasion, the node transits to IDLE state. In addition, on average, it takes half of the backoff window size steps for the counter to expire. The probability of above event is obtained as $p_{nk}(\bar{\sigma} \frac{W_0}{2})$ (here $\bar{\sigma}$ is the average slot time between successive backoff counter decrements. This parameter will be defined later). On the other hand, the node in question meets aforementioned backoff stages via two different ways in the diagram: first from state S_0 when the queue is empty (i.e. $q p_{S\_0}$) and second, when the packet is in FirstTR state and it sends a packet successfully without any arriving packets. Note that successful

transmission takes place with probability $(1 - c)$ and $p_{nk}(T_s)$ denotes the probability of no packet arrival during a successful transmission. So we can formulate the probability of IDLE state as follows:

$$p_{IDLE} = (q p_{S_0} + p_{FirstTR}(1 - c) p_{nk}(T_s)) p_{nk}(\bar{\sigma} \frac{W_0}{2}) \qquad (6)$$

State S_0 is accessed from states: $(i, 0)$ $(0 \leq i < m)$ with probability $1 - c$ (successful transmission) as well as state $(m, 0)$ with probability 1 (regardless of successful or failed transmission). Hence,

$$p_{S_0} = (1 - c) \sum_{i=0}^{m-1} p_{i,0} + 1 \times p_{m,0} \qquad (7)$$

Finally, $p_{0,0}$ can be found by normalizing it as follows:

$$1 = \sum_{w=0}^{W_0} p_{0',w} + \sum_{i=0}^{m} \sum_{w=0}^{W_i-1} p_{i,w} + p_{IDLE} + p_{FirstTR} \qquad (8)$$

In above equation, $\bar{\sigma}$ is the average slot time between successive backoff counter decrements. Based on the different conditions of the channel, given that the node is in backoff state, $\bar{\sigma}$ can be written as:

$$\bar{\sigma} = p_{tr}(n-1)(p_s(n-1)(T_s + \sigma) + (1 - p_s(n-1))(T_c + \sigma)) + (1 - p_{tr}(n-1))\sigma \qquad (9)$$

3.1 Single hop network analysis

For single hop ad hoc networks, all the nodes can hear each other. So the probability that a given node A collides with any of the other nodes is equal to the probability that at least one of the other $n - 1$ nodes transmits a packet conditioned that A already has a packet to transmit. This fact is implicitly considered in the Markov chain shown in Fig. 1 (i.e., collision is considered in the states where the node has a packet to transmit). Therefore, we have:

$$c = p_{tr}(n-1) = 1 - (1 - \tau)^{n-1} \qquad (10)$$

In the light of discussion so far, the transmission probability of a given node τ is given by:

$$\tau = \sum_{i=0}^{m} p_{i,0} + p_{FirstTR}$$
$$= \sum_{i=0}^{m} p_{i,0} + P_{IDLE}(1 - p_{tr}(n-1))(1 - p_{nk}(\sigma)) \qquad (11)$$

3.1.1 Analysis of utilization in single hop network

Consider a single hop ad hoc network comprising of n nodes with each node having the packet arrival rate of λ_g packets per second. Let U denote the normalized channel utilization, i.e. the fraction of time that the channel is used

to successfully transmit user data. The probability of failure, c, is calculated as given above. This probability in single hop condition is equal to the probability of at least one of the $n - 1$ nodes (except current node) having a packet to transmit in current slot, given that the current node transmits in same slot. The probability that at least one node transmits in an idle slot is:

$$p_{tr}(n) = 1 - (1 - \tau)^n \qquad (12)$$

and the probability that a packet is transmitted successfully is given as:

$$p_s(n) = \frac{n\tau(1 - \tau)^{n-1}}{1 - (1 - \tau)^n} \qquad (13)$$

Knowing that the average duration of payload is A_p, we have:

$$U = \frac{p_{tr}(n) p_s(n) A_p}{(1 - p_{tr}(n))\sigma + p_{tr}(n) p_s(n) T_s + p_{tr}(n)(1 - p_s(n)) T_c} \qquad (14)$$

In the above equation, the numerator is the mean size of useful load (user's payload) per slot and the denominator is the average slot time. It is noteworthy to mention that the above analysis of utilization has been extracted from the approach in Shabdiz and Subramanian [3].

3.1.2 Analysis of delay in single hop network

In a wireless ad hoc network, the backoff service time, s_b, is the time duration spent in backoff states before a packet is transmitted successfully or dropped due to the maximum retransmission constraint. One can obtain the average backoff service time, $E[s_b]$, by determining the time spent in the backoff states conditioned that a packet is successfully transmitted after i collisions $(i < m)$, or dropped after m collisions. Remember that after m' collisions, the backoff window size is constant between $m' \leq i \leq m$ and with the occurrence of the $m + 1$th collision the packet will be discarded.

To compute $E[s_b]$, we must compute the average time spent by a packet in each backoff stage. For the values of $i \leq m$ we suppose that the packet collides $i - 1$ times and transmits successfully at the last retry. Here c and $(1 - c)$ denote the probability of collision and probability of successful transmission, respectively. Moreover, iT_c denotes amount of time spent for each collision occasion. The successful transmission would take place within the period of T_s. Also for each stage, in average, the backoff counter expires after half of maximum window size decrement. Furthermore, remember from Sect. 3 that $\bar{\sigma}$ is the average slot time. So the average time spent between successive backoff counter decrements can be computed as

$\bar{\sigma}\frac{W_k-1}{2}$ ($W_k - 1$ is the maximum window size of stage k). Note that in step $m + 1$, the packet will be dropped after the $m + 1$th collision. In the light of above discussions, we obtain $E[s_b]$ as follows:

$$
\begin{aligned}
E[s_b] &= \sum_{i=0}^{m} c^i(1-c)\left(iT_c + \bar{\sigma}\sum_{j=0}^{i}\frac{W_j-1}{2} + T_s\right) + c^{m+1}T_c \\
&= \sum_{i=0}^{m-1} c^i(1-c)\left(iT_c + \bar{\sigma}\sum_{j=0}^{i}\frac{W_j-1}{2} + T_s\right) \\
&\quad + c^m\left((m+c)T_c + \bar{\sigma}\sum_{k=0}^{m}\frac{W_k-1}{2} + (1-c)T_s\right)
\end{aligned}
\tag{15}
$$

On further simplifying Eq. 16, we obtain:

$$
E[s_b] = (1-c^{m+1})\left(T_s + \frac{cT_c - \bar{\sigma}/2}{1-c}\right) + \bar{\sigma}/2\left(\sum_{i=0}^{m}W_ic^i\right)
\tag{16}
$$

The average service time in a single hop ad hoc network, A_s, consists of the average backoff service time (when the packet at its arrival finds the queue is nonempty) and the service time of packets when the node has no packet in the buffer. Therefore, A_s has to be conditioned on q as follows:

$$
\begin{aligned}
A_s &= (1-q)E[s_b] + q(p_{tr}(n-1)(c(T_c + E[s_b]) + (1-c)T_s) \\
&\quad + (1-p_{tr}(n-1))E[s_b])
\end{aligned}
\tag{17}
$$

Remember from previous section that q is the probability of the transmit queue being empty upon serving the current packet in backoff stage. This occurs when the node enters backoff stage with exactly one packet as well as having no more packets arriving while the current packet transmission is in progress. In the wake of statistically independence of these two events, we have:

leaves state n. Hence, for π_0 and π_1 which are respectively proportional in time when the system is empty and has exactly one packet, we get following equation:

$$
\lambda_g\pi_0 = E[s_b]\pi_1
\tag{19}
$$

To compute π_0, we note that the average number of service periods in a system with service time $E[s_b]$ and entrance rate λ_g is of form $\lambda_g E[s_b]$. However, as the left-hand side equals $1 - \pi_0$, we have:

$$
\pi_0 = \begin{cases} 1 - \lambda_g E[s_b] & E[s_b] \le \frac{1}{\lambda_g} \\ 0 & \text{otherwise} \end{cases}
\tag{20}
$$

From all the above equations and some algebraic simplifications, we have:

$$
q = \begin{cases} \lambda_g(\frac{1}{E[s_b]} - \lambda_g)e^{-\lambda_g E[s_b]} & E[s_b] \le \frac{1}{\lambda_g} \\ 0 & \text{otherwise} \end{cases}
\tag{21}
$$

Under light load conditions, the probability of entering backoff with one packet is almost one. It happens because the backoff states are hardly ever visited, and the probability of entering backoff with more than one packet is expected to be very small. For heavy loads where both, λ_g and $E[s_b]$ are large, the probabilities of q and entering backoff with one packet are almost zero. Suppose T_w is the average time that a given packet waits for another packet to be served. This time consists of random backoff period, the collision period, and the successful transmission period. For calculating T_w using M/G/1, we must obtain the probability distribution of T_w and its two moments. The distribution of T_w can be written by conditioning it on two events: (1) the channel and the node are idle on packet arrival, and (2) either, the channel or the node is busy when the new packet arrives. The conditional distribution of T_w based on the conditions above can be written as follows:

$$
\text{Prob}(T_\omega = T) = \begin{cases} q(1-p_{tr}(n-1))(1-c) & \text{if} \quad T = T_s \\ c^{m+1} & \text{if} \quad T = (m+1)T_c + \bar{\sigma}\sum_{j=0}^{m-1}\frac{W_j-1}{2} \\ 0 & \text{if} \quad T = T_s + iT_c + \bar{\sigma}\sum_{j=0}^{i-1}\frac{W_j-1}{2} \quad 0 < i \le m \end{cases}
\tag{22}
$$

$$
q = \pi_1 \times \frac{(\lambda_g t)^0}{0!}e^{-\lambda_g t} \qquad (t = E[s_b])
\tag{18}
$$

where π_1 and $E[s_b]$, respectively are the probability of having only one packet in the buffer and the average backoff service time which is obtained from Eq. 17. Considering the long run backoff procedure, the rate at which the process enters some particular state n equals the rate at which it

We consider the packet drop after m retransmission as part of the time it should wait. But in service time, we condition the transmission on the constraint of the last successful transmission. The task is to determine the first and second moments of T_w. Using the definition of the probability generation function Kleinrock [14], the Z-transform of the waiting time with b_s denoting the backoff stage can be written as:

$$G_{b_s}(z) = \sum_{i=0}^{\infty} z^i \Pr[b_s = i]$$

$$= z^0 \Pr[b_s = 0] + \sum_{i=1}^{m} z^i \Pr[b_s = i]$$

$$+ z^{m+1} \Pr[b_s = m + 1]$$

$$= T_s(1-c)q(1 - p_{tr}(n-1))$$

$$+ \sum_{i=0}^{m} \left(T_s + iT_c + \bar{\sigma} \sum_{j=0}^{i-1} \frac{W_j - 1}{2} \right) c^i(1-c)z^i$$

$$+ \left((m+1)T_c + \bar{\sigma} \sum_{j=0}^{m-1} \frac{W_j - 1}{2} \right) c^{m+1} z^{m+1}$$

In above equation, the maximum value of i is $m + 1$, which indicates the maximum retransmission attempts per packet. For the first moment of T_w, we have:

$$E(T_w) = \frac{dG_{b_s}(z)}{dz} \Big|_{z=1}$$

Which further can be written as:

$$E(T_w) = T_s(1-c)q(1 - p_{tr}(n-1))$$

$$+ (1-c^m)\left(cT_s + \frac{c(1+c)T_c - c\frac{\bar{\sigma}}{2}}{1-c} \right) + \frac{\bar{\sigma}}{2} c \sum_{j=0}^{m-1} W_j c^j$$
(23)

On the other hand, the second moment of T_w can be obtained as follows:

$$E[T_w^2] = \frac{d^2 G_{b_s}(z)}{dz^2} \Big|_{z=1} + E(T_w)$$

Which further yields (the detailed analysis is described in Appendix B):

$$E[T_w^2] = T_s^2(1-c)q(1 - p_{tr}(n-1))$$

$$+ T_s^2(c - c^{m+1}) + 2T_s T_c \left(-mc^{m+1} - \frac{c(1-c^m)}{1-c} \right)$$

$$+ T_c^2 \left(\frac{2mc^{m+1}}{1-c} + (2m+1)c^{m+1} + \frac{c(1+c)(1-c^m)}{(1-c)^2} \right)$$

$$+ \frac{-\bar{\sigma}}{4} \sum_{j=0}^{m-1} (W_j - 1)^2 c^{j+1} + T_s \bar{\sigma} \sum_{j=0}^{m-1} (W_j - 1)c^{j+1}$$

$$+ (T_c - T_s)\bar{\sigma} c^m \sum_{j=0}^{m-1} W_j - 1 + T_c \bar{\sigma} \sum_{j=0}^{m-1} (W_j - 1)(j+1)c^{j+1}$$

$$+ T_c \bar{\sigma} \sum_{i=j+2}^{m} \sum_{j=0}^{m-2} (W_j - 1)c^j$$
(24)

Having the second moment of T_w, the average waiting time in the M/G/1 queue (A_w) can be obtained using Pollaczek-Khinchin Bertsekas and Gallager [15] as given below:

$$A_w = \frac{\lambda_g E[T_W^2]}{2(1 - \lambda_g E[T_W])}$$
(25)

The average single hop delay (A_d) in the network is the sum of the average service time and the average queuing delay. Hence, we have:

$$A_d = A_s + A_w$$
(26)

4 Multi hop network analysis

In multi hop ad hoc network, in contrast to the single hop scenarios, all nodes are not visible to each other. According to DCF, a node can initiate a transmission only if it senses the medium as being idle for a time interval greater than a DIFS. Unlike the wired networks (with CSMA and collision detection support), in a wireless network, collision detection is not possible. Hence, an acknowledgement (ACK) frame is used to notify the sender that the transmission has been successful. The discussed transmission mechanism does not protect the nodes from the hidden terminal problem which is depicted in Fig. 2(a). According to this problem, given station A is considered to be hidden from another station B in the same area of coverage of the receiver if the transmission coverage of the transceivers at A and B do not overlap. In these situations, if both of these nodes initiate a transmission with a third node located in the interference area of both A and B, the hidden terminal problem would occur. In order to alleviate the problem of hidden nodes in ad hoc networks, the IEEE 802.11 uses virtual carrier sense mechanism, which is based on the exchange of the two control messages called (Request to send) RTS and (Clear to send) CTS.

As another problem in multi hop ad hoc networks, exposed node problem occurs when a node unnecessarily defers its transmission without any collision occurrence at the destination. Consider Fig. 2(b). B is able to communicate with A and C. C can communicate with B and D, but it lies beyond the coverage of A. During the communication from B to A, C senses the medium busy and defers the ongoing transmission to D, despite the fact that this transmission would not result in a collision at A and D. In this case, C is exposed to B, which degrades the performance of the network. In what follows, we discuss the delay analysis of multi-hop wireless network with and without considering the exposed terminal problem. Let $G(N, \varepsilon)$ be the graph that represents a wireless ad hoc network with node set N and edge set ε, with each edge between nodes indicating the direct communication
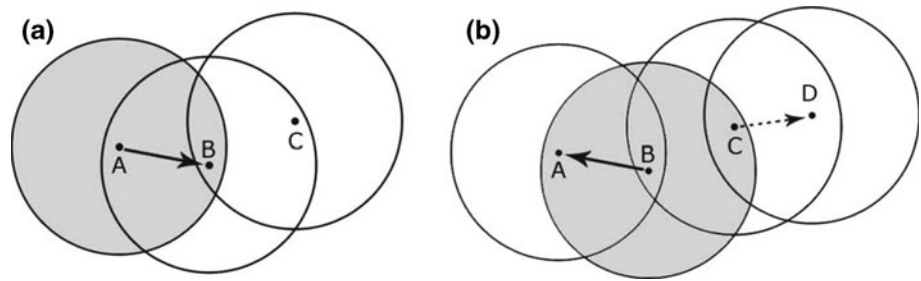
between the nodes; that is, the nodes with Euclidean distance less than transmission range of each other ($r(n)$) are connected by an edge. Due to the uncertainty in the node locations, the edge set ε changes randomly. The parameter ε can also be interpreted as uncertainty in node locations due to slow speed movements. In order to study topological properties, $G(N, \varepsilon)$ is best modeled as a geometric random graph Kumar (2004). Let nodes be distributed uniformly in the square operational area $[0, z]^2$. In this network with the average node density $\frac{N}{z^2}$, the threshold value of $r(N)$ for large values of n is $r(N) = \sqrt{(\ln n)/n}$ Kumar (2004). Also, the average distance $\bar{r}_{i,j}$ between nodes i and j is given by $\bar{r}_{i,j} = 2/3r(N)$ (see Appendix 2).

Now we proceed with finding the probability of failure c, given that the node transmits a packet. Considering that the node S initiates communication with the destination B by transmitting RTS in a given slot. This transmission will be successful if all the following conditions hold:

1.1 No RTS message transmission occurs in the interference area of S.

1.2 No CTS message transmission is in the interference area of S in the given slot (the source of any probable CTS is out of the interference area of S. It is due to the fact that we already know that during previous time slot, there was no RTS transmission in the interference area of S; otherwise the ongoing RTS transmission would not happen. So the initiator of the CTS must be located outside the interference region of S).

1.3 There exists no message transmission in the hidden area of S. The aforementioned conditions assure that the RTS message arrives successfully at the destination, after which the destination transmits the CTS to S. To accomplish this task successfully, the following conditions must hold simultaneously:

1.4 None of the neighbors of B in the hidden area of S should be receiving data. It means that there should not be a successful transmission between a node located in the hidden area of the node B and any neighbors of B located in the hidden area of S.

1.5 There should be no RTS message transmission in the hidden area of S.

1.6 There should be no CTS message transmission in the hidden area of S (the source of this CTS is beyond the interference area of B; otherwise the ongoing CTS message would not take place).

For the first event, none of the neighbors of S must transmit any RTS message in the given slot. Considering a particular node x, the average number of nodes located within the coverage disk of x (other than itself) is given by $\frac{N}{z^2}\pi r^2(n) - 1$. Remember from the Sect. 3.1 that τ is the probability of transmission in an arbitrary slot. So, the probability of no message transmission in the vicinity of a particular node is given by:

$$(1 - \tau)^{\frac{N}{z^2}\pi r^2(N) - 1}$$

Regarding the second event, we emphasize on the probability of no CTS transmission in the given slot, conditioned that there was no RTS transmission in the interference area of S. A CTS transmission in the interference area of S cannot be a response to a RTS message previously originated in the interference area of S, otherwise the channel would be busy and first event would not have occurred. So, we are interested in cases where a node in the interference area of S sends a CTS message in response of the RTS message originated from somewhere outside the sensing area of S. Therefore, node S has to be located in the hidden area of the node which transmits the first RTS message. The probability of CTS message transmission in the given slot can be approximated by the probability of successfully receiving an RTS, such that the source of the RTS is located in the hidden area of S. Assume n_{Hid} to be the average number of nodes in the hidden area of node S. Thus, n_{Hid} hidden nodes of S with probability τ request to transmit and this request is transmitted successfully to the destination with probability $(1 - c)$. Now, the probability of no node in the hidden area that successfully transmits during a given time slot is given by:

$$(1 - \tau(1 - c))^{n_{\text{Hid}}}$$

Owing to the fact that every node in the interference area of S could receive this request during the same slot in which S transmits, the probability of the second event can be formulated as:

$$((1 - \tau(1 - c))^{n_{\text{Hid}})\frac{N}{z^2}\pi r^2(N) - 1}$$

Concerning Fig. 3, we need to calculate n_{Hid}. To do so, we need to calculate the area AECF.

$$\text{AECF} = \pi r^2(N) - \text{ABCE}$$

$$\text{ABCE} = 2r^2(N)acos\left(\frac{d}{2r(N)}\right) - (\text{ASCD})$$

$$\text{ASCD} = \frac{1}{2}d2r(N)\sin\left[acos\left(\frac{d}{2r(N)}\right)\right]$$

Substituting the above equations, we get:

$$\text{ABCE} = 2r^2(N)acos\left(\frac{d}{2r(N)}\right) \\ - dr(N)\sin\left(acos\left(\frac{d}{2r(N)}\right)\right) \quad (27)$$

and

$$\text{AECF} = \pi r^2(N) - 2r^2(N)acos\left(\frac{d}{2r(N)}\right) \\ + dr(N)\sin\left[acos\left(\frac{d}{2r(N)}\right)\right] \quad (28)$$

For the average case where $d = \bar{r}_{i,j} = 2/3r(N)$, the above equation becomes $AECF = 1.3082r^2(N)$ and n_{Hid} becomes:

$$n_{\text{Hid}} = \frac{N}{z^2}1.3082r^2(N). \quad (29)$$

In addition, $ABCE = 1.8338r^2(N)$ and if we refer to the average number of nodes in this area as n_c, then we have:

$$n_c = \frac{N}{z^2}1.8338r^2(N) \quad (30)$$

The third event indicates the probability of no message transmission in the hidden area during the time slot. In average, there are n_{Hid} nodes in the hidden area of S, thus the probability of no message transmission in the hidden area would be:

$$(1 - \tau)^{n_{\text{Hid}}}$$

After above three events, the RTS message is sent successfully. Now it is time for the destination node B to
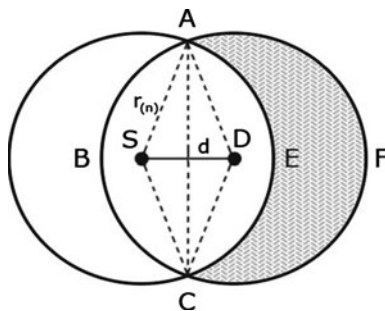

Fig. 3 The hidden area of the given node S

transmit the CTS. Events 4, 5 and 6 should be checked in a different time slot rather than the current time slot (of duration σ). We refer to this time as the vulnerable period t_v. For the case of no hidden stations we have $t_v = \sigma$. But in multi hop ad hoc networks with hidden terminals, the vulnerable period is equal to Marsic [4]:

$$t_v = \text{RTS} + \text{SIFS} + \sigma \quad (31)$$

Event 4 indicates the probability that none of neighbors of destination node B that are hidden to the source node S are involved in a successful communication. The reason why we only consider the hidden terminals of S in this event is that if one of nodes in the sensing area of S is involved in a communication, then the channel would be busy and the ongoing communication would not be initiated. The probability that a neighbor marks the channel busy on successfully receiving data is:

$$\frac{\tau(1 - c)(T_s - t_v)}{\bar{\sigma}} \quad (32)$$

where $\frac{T_s - t_v}{\bar{\sigma}}$ is the fraction of time that the channel is busy during a successful reception. Considering node x in the sensing area of B, we want to calculate the probability of having a busy channel because of successfully receiving data from a node in the hidden area of B. We have in average $\frac{N}{z^2}\pi r^2(n)$ numbers of nodes in the vicinity of node B. Thus, regardless of which node actually receives data, this probability is obtained by:

$$\left[\frac{1}{\frac{N}{z^2}\pi r^2(N)}\frac{\tau(1 - c)(T_s - t_v)}{\bar{\sigma}}\right]^{n_{\text{Hid}}}$$

On the other hand, node x is selected from the sensing area of node B, the probability that none of neighbors of B in the hidden area of S are receiving data over t_v is equal to:

$$\left[1 - (\frac{\tau(1 - c)(T_s - t_v)}{\bar{\sigma}\frac{N}{z^2}\pi r^2(N)})^{n_{\text{Hid}}}\right]^{n_{\text{Hid}}\frac{t_v}{\sigma}}$$

The probability of event 5, i.e. the probability of no RTS transmission in the hidden area of S over duration t_v is given as:

$$(1 - \tau)^{n_{\text{Hid}}\frac{t_v}{\sigma}}$$

Finally, event 6 is obtained in a manner similar to event 2. This event regards the probability of having no CTS transmission by a node in the hidden area of S over the vulnerable period t_v, given that there was no RTS transmission in that area. A CTS transmission in the hidden area cannot be a response to an RTS transmission that originated in the hidden area, otherwise the destination channel would be busy. Therefore, using the same analogy as in the calculation of event 2, the fraction of nodes with

an idle channel in the hidden area has to be considered. The probability of CTS message transmission in the given slot can be approximated by the probability of successfully receiving an RTS, and the source of the RTS is located in the hidden area of B. The probability that none of the hidden nodes of S transmit CTS over vulnerable period t_v is formulated as:

$$((1-\tau(1-c))^{n_{\text{Hid}}})^{n_{\text{Hid}}\frac{t_v}{\sigma}}$$

After all above events happen, the ongoing transmission would be successful. Therefore, taking into consideration the above mentioned conditions of independent events leads us to the probability of failure in multi-hop network:

$$c = 1 - (1-\tau)^{(\frac{N}{z^2}\pi r^2(N)-1)+n_{\text{Hid}}(1+\frac{t_v}{\sigma})}$$
$$\times ((1-\tau(1-c))^{n_{\text{Hid}}})^{(\frac{N}{z^2}\pi r^2(N)-1)+n_{\text{Hid}}\frac{t_v}{\sigma}}$$
$$\times \left[1-(\frac{\tau(1-c)(T_s-t_v)}{\bar{\sigma}\frac{N}{z^2}\pi r^2(N)})^{n_{\text{Hid}}}\right]^{n_{\text{Hid}}\frac{t_v}{\sigma}}$$

In above equations, we do not consider the exposed terminal problem. In order to derive the media access delay in absence of the exposed terminal problem, we assume that the IEEE 802.11 standard includes some algorithms where the nodes are aware of when to transmit and when not to, during another transmission in progress. Following this line of thought, in a similar way to the previous section, given that the node S initiates communication with the destination B by transmitting RTS in a given slot, the transmission will be successful if all the following conditions hold:

2.1 No RTS message transmission occurs in the common interference area of nodes S and B (here the main difference with the event 1.1 is that with a proper strategy for the exposed node problem in IEEE standard, nodes in the interference area of S that are out of the transmission range of B would not prevent from transmission with some node out of the interference area of B).

2.2 No CTS message transmission occurs in the common interference area of S and B in the given slot (the source of this CTS is out of the common interference area, i.e. there is no RTS transmission in the common interference area, otherwise the ongoing transmission would not happen).

2.3 There exists no message transmission in the hidden area of S. Now the destination transmits the CTS to S. To accomplish this task successfully during the vulnerable period, the following conditions must hold simultaneously:

2.4 None of the neighbors of B in the hidden area of S should be receiving data. It means that there should not be a successful transmission between a node

located in the hidden area of the node B and one of the neighbors of B which is located in the hidden area of S.

2.5 There should be no RTS message transmission in the hidden area of S.

2.6 There should be no CTS message transmission in the hidden area of S (the source of this CTS is beyond the interference area of B, otherwise the ongoing CTS message would not be sent).

2.7 None of the neighbors of S located out of the common interference area of S and B(the area of S that is hidden to B) should be transmitting an RTS.

2.8 There is no CTS message transmission in the interference area of S that is hidden to B(the source of this CTS is out of common interference area of nodes S and B, otherwise the ongoing CTS message would not be sent).

For the first event, none of the nodes in the common interference area of nodes S and B must transmit RTS in the given slot. Knowing τ is the probability of transmission in an arbitrary slot and n_c indicates the average number of nodes in common area, the probability of no message transmission occurrence in common area of nodes S and B (event 2.1) is given by.

$$(1-\tau)^{n_c-1}$$

Event 2.2 indicates that there should be no CTS message in the common interference area of S and B, given that there was no RTS transmission in that area. A CTS transmission in this area cannot be a response to an RTS in which previously originated in the common area, otherwise the channel would have been busy and event 2.1 would not have occurred. The probability of CTS message transmission in the given slot can be approximated by the probability of successfully receiving an RTS, and the source of the RTS is located out of the common interference area of nodes S and B. The average number of nodes located out of shared interference area is given by $\frac{N}{z^2}\pi r^2(n) - n_c$. These nodes request to transmit with probability τ and this request is transmitted successfully to the destination with probability $(1-c)$. Owing to the fact that every node in the common interference area of S and B could receive this request at the same slot that S transmits, similarly to the event 1.2, the probability of event 2.2 can be formulated as

$$((1-\tau(1-c))^{\frac{N}{z^2}\pi r^2(N)-n_c})^{n_c}$$

The probability of events 2.3 to 2.6 are the same as that of events 1.3 to 1.6. Event 2.7 is equal to the probability of no RTS transmission in a portion of interference area of S which is hidden to the node B. This event happens over the vulnerable period t_v. Again, considering τ as the probability of sending packet in current slot and $\frac{N}{z^2}\pi r^2(n)$ $-n_c$ as the average number of nodes adjacent to S and

beyond the interference area of B, event 2.7 can be written as:

$$(1 - \tau)^{(\frac{N}{z^2}\pi r^2(N) - n_c)\frac{t_v}{\sigma}}$$

Eventually, event 2.8 is obtained correspondingly to event 2.2. This event regards the probability of having no CTS transmission by a node in the interference area of S and hidden to B, given that there was no RTS transmission in that area. For the sake of conciseness, this probability over the duration t_v is as:

$$((1 - \tau(1 - c))^{\frac{N}{z^2}\pi r^2(N) - n_c})^{n_c\frac{t_v}{\sigma}}$$

Finally, note that the above defined events are independent of each other. Therefore, the probability of failure in multi-hop network with eliminating exposed terminal effect can be rearranged as follows:

$$c = 1 - (1 - \tau)^{(n_c - 1) + n_{Hid}(1 + \frac{t_v}{\sigma}) + (\frac{N}{z^2}\pi r^2(n) - n_c)\frac{t_v}{\sigma}}$$
$$\times ((1 - \tau(1 - c))^{\frac{N}{z^2}\pi r^2(N) - n_c})^{n_c(1 + \frac{t_v}{\sigma})}$$
$$\times ((1 - \tau(1 - c))^{n_{Hid}})^{n_{Hid}\frac{t_v}{\sigma}}$$
$$\times \left[1 - \left(\frac{\tau(1 - c)(T_s - t_v)}{\bar{\sigma}\frac{N}{z^2}\pi r^2(N)} \right)^{n_{Hid}} \right]^{n_{Hid}\frac{t_v}{\sigma}}$$

On the other hand, to estimate multi-hop delay we must study network traffic patterns and routing schemes. The traffic pattern affects mainly the mean message distance, $\bar{d}$ which is the expected number of hops (i.e. the number of intermediate nodes) that a message makes to reach its destination. The mean message distance is generally given by

$$\bar{d} = \sum_{i=1}^{D} ip_i \tag{33}$$

Where p_i is the probability of a message crossing i channels before reaching its destination and D is the diameter of the network. The diameter is the maximum value (in hops) of the minimum distance between any pair of nodes. Different choices of p_i lead to different distributions for message destinations and consequently, to various mean message distances. The following analysis uses the decreasing probability routing distribution defined in Reed and Fujitomo [16] as a model of communication locality (it is worth noting that the modeling approach presented here can be equally applied to the other models discussed in Reed and Fujitomo [16]). In this model, the probability of sending a message to a particular destination node, i hops away, decreases with the distance i. For an ad hoc network, the probability $p_i(1 < i < n)$ can be defined as:

$$p_i = \theta(\alpha, n)\alpha^i \tag{34}$$

where α is between 0 and 1 leading to varying degrees of communication locality. As α approaches zero, the degree of locality increases, while as α approaches 1, the traffic becomes more uniform. The factor $\theta(\alpha, n)$ is a normalizing constant, and is chosen such that:

$$\theta(\alpha, n) \sum_{i=1}^{n} \alpha^i = 1 \tag{35}$$

From the above equation, we can easily determine $\theta(\alpha, n)$. Substituting the expressions of $\theta(\alpha, n)$ in equations Eqs. 33 and 34 yield the probability p_i and the mean message distance, $\bar{d}$ as:

$$p_i = \frac{(\alpha - 1)\alpha^{i-1}}{\alpha^n - 1} \tag{36}$$

$$\bar{d} = \frac{(na - n - 1)\alpha^n + 1}{(\alpha - 1)(\alpha^n - 1)} \tag{37}$$

In an ad hoc network, each node could be a source, destination and/or relay of packets. We assume that packets are generated in each node with the rate of λ_g packets per second. Given the number of nodes in the network N and mean message distance $\bar{d}$, the traffic of the network λ_n is as follows:

$$\lambda_n = N \cdot \lambda_g \cdot \bar{d} \tag{38}$$

Having the probability of collision, c and the traffic of the network λ_n, we can obtain average packet delay for 1-hop in multi-hop ad hoc networks with the same method as used in single-hop analysis (from Eq. 26). Recalling A_d as the average delay at each hop, the average end-to-end delay equals the product of the average number of hops traversed by a packet ($\bar{d}$) and the average delay at each node A_d. Hence we have:

Latency = $\bar{d}A_d$.

5 Simulation results

In order to verify the accuracy of our proposed analytic model for single-hop and multi-hop networks, we have considered a variety of scenarios using Pythagor simulator (2004). This simulator is an open C++ tool for IEEE 802.11 a/b/g networks. Also, we have modified this platform to include multi-hop scenarios, routing schema and the exposed terminal problem. This simulator performs a detailed implementation of the MAC mechanism for all extensions of IEEE 802.11 standard (basic mode, RTS/CTS and etc). A variety of statistic metrics for network performance evaluation such as throughput (in bits/s and packets/s), utilization, media access delay, queuing delay, total packet delay, packet queue length, and packet

retransmission attempts have been implemented in this simulator. Moreover, ease in configuring each node parameter (using network configuration files) permits us to define various scenarios by defining input system parameters. Regardless of the major advantages of Pythagor (2004), we have slightly modified the software to include multi hop transmission scenarios, defining the final destination of packets at the source node, routing schema and the exposed terminal issue.

Our proposed model holds for any input traffic model, as long as the parameters λ_g and q are defined in closed form expressions. Moreover, for the packet payload size, an exponential distribution with a mean value of P bits is chosen. It is known that the exponential distribution is appropriate for the approximation of the packet size distribution in standard IEEE 802.11 networks Bianchi [6]. We have implemented the buffer of each node as a linked list structure in which new packets are inserted at the end of the list. Thus, the buffer space can be regarded as infinite.

The initial network size is 200 nodes that are deployed in a uniform distribution in a network of area $1,500 \times 1,500$ m. The minimum backoff window size (W) is 32 and m (maximum number of retransmissions) is equal to 5. A summary of input system parameters used during the validation experiments is shown in Table 1. The analytical model has been validated through a batch means discrete-event simulator for each simulation experiment, 25 batches of messages were delivered for collecting the statistics of interest; each batch consisted of 10,000 messages. Statistics gathering was inhibited for the first batch to avoid distortions due to the initial warm-up conditions. We adopted 95 per cent confidence level to make sure that on average the confidence interval calculated; using t-student distribution and standard error, is expected to contain the true value around 95 per cent of the time.

Table 1 The system parameters used in the simulation process

Parameter	Value
MAC header	272 bits
Physical header (PHY)	96 bits
ACK	112 bits + PHY header
RTS	160 bits + PHY header
CTS	112 bits + PHY header
SIFS	10 μs
DIFS	50 μs
Minimum contention window (W)	31 slot
Slot duration	20 μs
Channel bit rate	5.5 Mbps
Timeout	300 μs
RTS threshold	1,024 bits
Packet payload (P)	10 kbit

It is well known that the accuracy of the model degrades as the network approaches the heavy traffic region. This is due to the approximations that have been made in the analysis to ease the model development. For example it is known that there is interdependency between the inter-arrival time and service time at intermediate network channels in a multi hopping network. However, calculating waiting time at a channel with the dependency assumption is far from trivial for a queueing network and it is still a open problem in the research community of queueing systems.

5.1 Single hop network simulation

In single hop networks, all nodes are in line of sight with each other. In other words, any two given nodes in the network are able to communicate directly. In the following validation experiments, the network utilization and access delay parameters are calculated for the single hop scenarios. Transmission rate for each node is considered based on IEEE 802.11g standard (54 mbps). Time duration of each simulation run is 30 s. This amount of time is sufficient for the system load to reach saturation condition (this case is verified by several simulation tests).

In the first single-hop scenario, we studied the case of 10 and 20 active nodes. The packet arrival rate was increased so that the system load reached saturation gradually. Figure 4 and Table 2 compares the medium access delay for the analytic results and the simulation results under different system load for this case. Figure 4 and Table 2 indicates that the medium access delay increases with increase in system load. When the system load reaches saturation, however, the medium access delay does not increase any further. This is mainly due to the fact that packets arrive so fast that the system cannot serve them. Larger number of active nodes results in greater collision probability and thus, the medium access delay become longer.
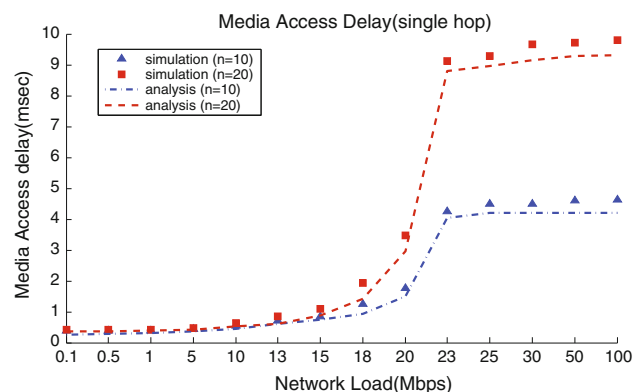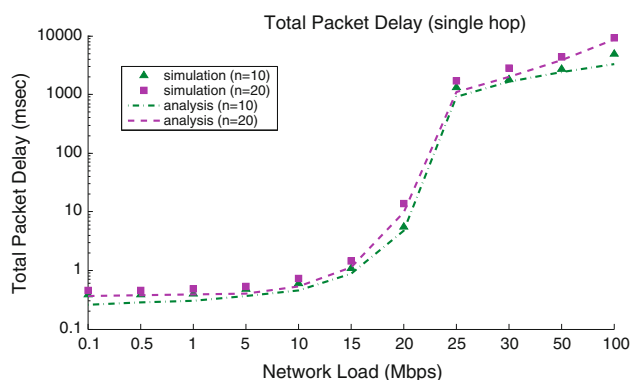

Fig. 4 Media access delay in single hop network

Table 2 Media access delay in single hop network

Network load (Mbps)	10 nodes		20 nodes	
	Difference	Relative error	Difference	Relative error
0.1	0.121	0.3192	0.0445	0.1076
0.5	0.099	0.2551	0.0405	0.0967
1	0.084	0.2142	0.0295	0.0693
5	0.088	0.1925	0.0775	0.1599
10	0.108	0.1908	0.1115	0.1749
15	0.106	0.1254	0.23	0.2089
20	0.275	0.1558	0.4985	0.1439
25	0.291	0.0648	0.308	0.0332
30	0.295	0.0654	0.5195	0.0537
50	0.403	0.0873	0.413	0.0425
100	0.418	0.0902	0.486	0.0496

Fig. 5 Total packet delay in single hop

**Table 3** Total packet delay in single hop network

Network load (Mbps)	10 nodes		20 nodes	
	Difference	Relative error	Difference	Relative error
0.1	0.132	0.3384	0.0815	0.1809
0.5	0.105	0.2664	0.085	0.1835
1	0.092	0.23	0.0875	0.1809
5	0.11	0.2296	0.128	0.2392
10	0.153	0.2504	0.1925	0.2679
15	0.19	0.1759	0.3415	0.2336
20	0.866	0.1545	3.94	0.2852
25	395.69	0.2999	583.44	0.3469
30	151.90	0.0844	806.00	0.2859
50	318.63	0.1157	430.50	0.0996
100	1702.26	0.3441	668.72	0.0718

Figure 5 and Table 3 shows the simulation and analytic results on the total packet delays for the case of 10 and 20 nodes. It can be conceived that when the network reaches

the saturation load, the average delay approaches to infinity. This is because when the system load is high, the queue becomes very long, and the queuing delay tends to infinity.

The relation between media access delay and network size in single hop network is investigated in the next experiment. It can be seen in Fig. 6 that the media access delay increases with the number of nodes in the network. This is mainly because the collision probability increases as a result of increased network size. As illustrated in this figure, the results from the simulation and analysis are reasonably close to each other.

In the next step, the network utilization factor is measured for the case of 10 and 20 nodes. At low packet arrival rate, the network utilization is around zero. This is because the network resources remain unused due to low data generation rate. When the packet arrival rate increases, network utilization also increases gradually till it reaches the balance condition. It happens when each node always has a packet to send (i.e. it is saturated). In such occasions, the network utilization will remain unchanged. As plotted in Fig. 7, the utilization factor for the case of 10 nodes is more than the situation when we have 20 nodes in the network. This is because the collision probability will
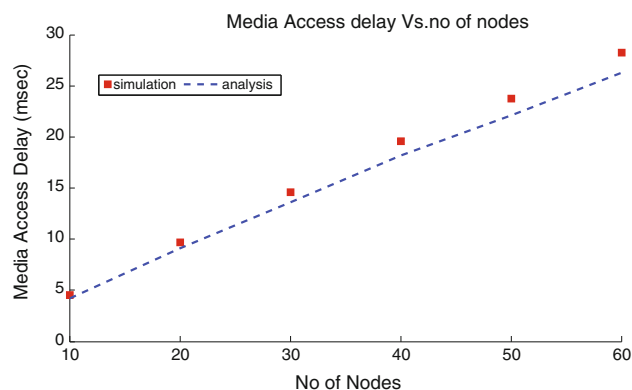
Fig. 6 Average medium access delay in saturation condition (30 mbps) for different network size
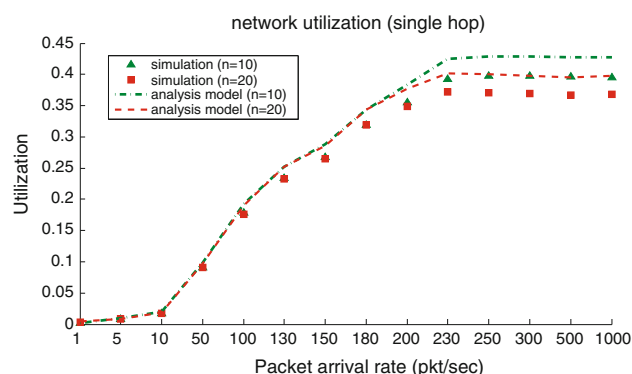
Fig. 7 Network utilization factor for different packet arrival rates and for the case of 10 and 20 nodes in a single hop network

increase due to the increased network size and consequently, result in degradedness in the system throughput.

5.2 Multi hop network simulation results

Network model and assumptions for multi hop networks are the same as those for a single hop network, except that in a multi hop network the destination of a packet might not be reached directly and other nodes can be used as relays to route the packets to the final destination. In the case of a multi-hop network, the packets arriving at a node are composed of newly generated packets and transit packets routed through the node. It is assumed that the total packet arrival rate at a node is known. The transit packet arrival rates can be calculated from the arrival rate of packets to the network, the traffic distribution and routing algorithm. In our analytical model, we have used the decreasing probability routing distribution (Reed and Fujitomo [16]) as the default routing mechanism. Moreover, for the analysis it is required that $\bar{d}$, the mean message distance to be known. Hence, in each experiment, with the definite number of nodes, details of nodes deployment, the routing circumstances and the rate of packet arrivals, we prepared the required parameters for the analytic model and then compared the results.

In order to validate our delay analysis for a multi-hop network with different loads, a 40 node network was considered, each node serving both, as a transmitter and a receiver. Figures 8 and 9 show the 1-hop to 5-hop medium access delay for the multi hop scenario. Packets generation rate increases from 1 to 30 packets per second. In this situation, the network load changes from 0.8 to 12 Mbps. The overall network traffic (λ_n) is obtained from the product of the values of packet generation rate of each node λ_g, network size N and mean message distance $\bar{d}$. For the coordination between analysis and validation experiments, for each case we have obtained the number of transmitted packets in each hop via simulation and then applied them in the analytical process. It is
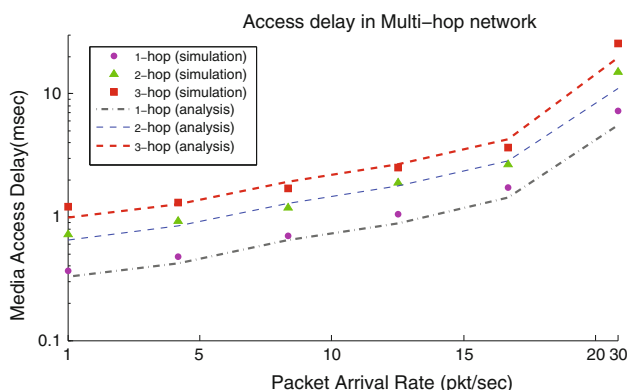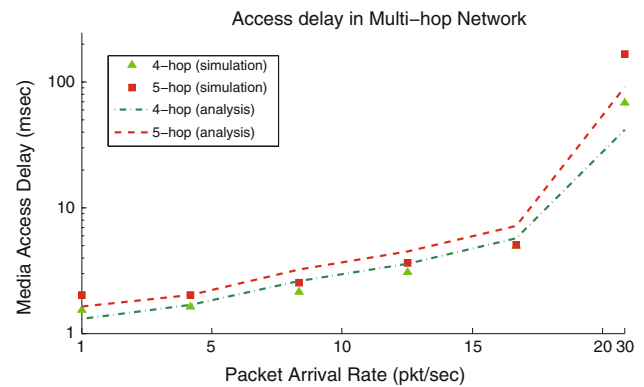

Fig. 9 Media access delay in multi hop network for 4, 5 scenarios

noteworthy to mention that in this experiment, we have considered only the effect of the hidden terminal problem in the performance evaluation. Separated from the effect of the increment in packet arrival rates, we can see the media access delay augments when the number of hops increases. This is due to the reason that as the load in the system increases, the collision probability in the network rises.

In our next experiment, we investigated the total packet delay from source to the destination. The mean message distance is taken to be 3 in this case. To obtain the end-to-end delay in a similar approach as the previous section, first we calculated total delay for different hop counts and then obtained the average total delay considering the amount of transmitted messages in each hop. Table 4 summarizes the results of the comparison between the analytical metrics and the simulation output. In all cases, the results show that the models are fairly accurate.

Figures 10, 11 and Table 5 illustrate the impact of the exposed terminal problem on the ad hoc network media access delay and total packet delay, respectively. In these experiments, the average media access delay and end-to-end delay in multi hop scenarios are obtained via two different cases. In first case, we concern real conditions, meaning existence of hidden and exposed nodes, and in second one, we assume that the IEEE 802.11 standard includes a sophisticated algorithm to eliminate exposed terminal problem. As obvious from Figs. 10 and 11, the average media access delay and average end-to-end delay

Fig. 8 Media access delay in multi hop network for 1, 2 and 3 hop

Table 4 Model validation for the multi hop end-to-end delay

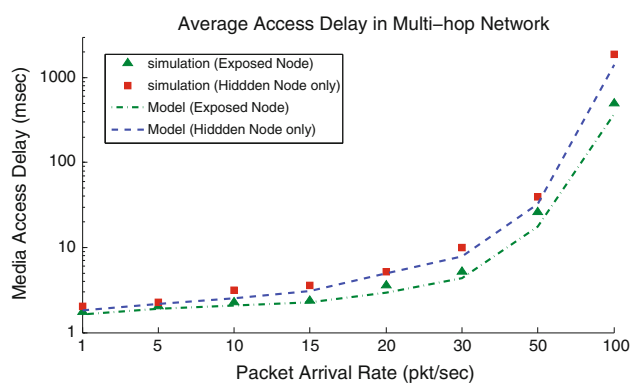| Network load (kbps)	End-to-end delay (ms) (simulation)	End-to-end delay (ms) (analysis)	Difference
800	1.26	1.01	19.84
4,000	2.37	2.17	8.43
8,000	3.63	3.29	9.36
16,000	6.94	6.65	4.17
24,000	855.01	795.44	6.96
40,000	2854.99	2487.67	12.86

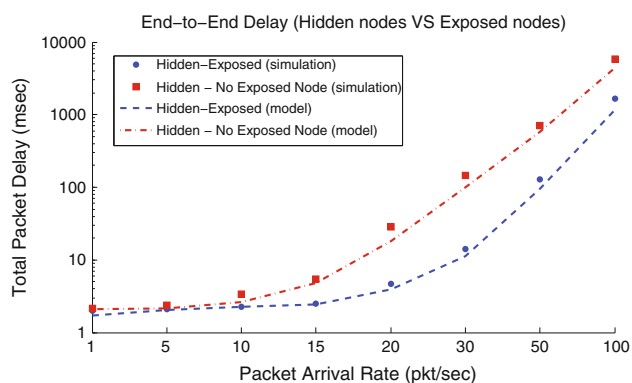**Fig. 10** Average medium access delay effect of exposed

Fig. 11 Average end-to-end delay effect of exposed

for the second case is less than the first one where exposed terminal problem remains unsolved in the network.

In the real-world case, where the both problems exist, hidden nodes transmit during the transmission of others and

exposed nodes defer their transmissions pointlessly. Consequently, as the network traffic increases, the collisions cause the media access delay and end-to-end delay to be longer than the case where the exposed terminal problem is treated. In this situation, the collisions are less and consequently result in lesser delay seen by each packet.

The figures reveal that the simulation results closely follow those predicted by the analytical model in the steady state regions, that is, under light and moderate traffic and when the network enters the heavy traffic region. However, the accuracy of the model degrades as the network approaches the saturation point in the heavy traffic region. This is due to the approximations that have been made in the analysis to ease the model development. For instance, treating the successive channels as independent of each other in the analysis. To overcome this dependence problem, we have used the well-known Kleinrocks independence approximation Kleinrock [14] that suggests an approximate solution for non-Jacksonian queueing networks. Nevertheless, it can be concluded that the model produces correct results in the steady state regions, and its simplicity makes it a practical and cost-effective evaluation tool.

6 Conclusions

In this paper, we have presented the analytical models for multi hop delay estimation in IEEE 802.11 wireless ad hoc networks under finite load conditions. For this objective, we obtained single hop delay using service time distribution function and its first and second moments. Moreover, considering effects of exposed terminals on network performance, we extended single hop analysis to the multi hop

Table 5 Average medium access delay and average end-to-end delay effect of exposed terminals

Packet arrival Rate (pkt/s)	1	5	10	15	20	30	50	100
Average medium access								
Hidden node—with exposed node								
Difference	0.133	0.118	0.199	0.104	0.624	0.833	8.526	126.09
Relative error	0.076	0.058	0.088	0.044	0.174	0.159	0.325	0.252
Hidden node—no exposed node								
Difference	0.181	0.109	0.606	0.523	0.291	2.127	6.968	439.1
Relative error	0.090	0.048	0.193	0.144	0.055	0.213	0.176	0.2351
Average end-to-end delay								
Hidden node—with exposed node								
Difference	0.263	0.032	0.020	0.061	0.728	3.091	33.75	502.5
Relative error	0.132	0.015	0.008	0.024	0.155	0.217	0.262	0.299
Hidden node—no exposed node								
Difference	0.057	0.204	0.706	0.616	10.59	43.26	131.1	1439.1
Relative error	0.026	0.086	0.211	0.113	0.371	0.299	0.183	0.245

scenario. In order to validate the accuracy of the analytic models, various simulation experiments have been taken into account. In all cases, the results of the simulations meet quite well with the analytical estimates. As the next step of this study, we aim at considering analytical models for other traffic patterns in wireless ad hoc networks, such as the model to deal with non-Poisson traffic.

Appendix 1: Table of notations

Notation	Description
A_p	Average duration of payload
A_s	Average service time
A_w	Average waiting time
A_d	Average End-to-end delay
s_b	Backoff service time
T_w	Average time that a packet waits for another one to be served
m	Maximum number of retransmission
m'	Maximum contention window
c	Probability of collision occurrence
q	Probability of empty queue after processing current packet
T_c	Collision slot duration
T_s	Successful slot duration
P_s	Probability of busy slot being a successful transmission
P_{tr}	Probability of sensing a busy slot
P_{nk}	Probability of generating no packet in a given period
U	Normalized channel utilization factor
τ	Probability of transmission in an Idle slot
W	Minimum contention window
W_i	ith contention window
σ	Slot duration
$\bar{\sigma}$	Average slot duration
$p_{i,j}$	Probability of being in the (i,j) state at the State transition diagram.
$p_{FirstTR}$	State of receiving new packet when both node and channel are idle
p_{IDLE}	State of the node when it has no packet in queue to serve
λ_g	Average packet arrival rate
$r(N)$	Communication range of node n
$\bar{r}_{i,j}$	Average distance between each two nodes in the network
n_{Hid}	Average number of nodes located in hidden area of a random node
n_c	Average number of node located in common area of two random adjacent nodes
t_v	Vulnerable period
$\bar{d}$	Mean message distance
p_i	Probability of a message crossing i channels to reach the destination
D	Diameter of the network
α	Indicating the degree of communication locality

Appendix continued

Notation	Description
$\theta\,(\alpha, n)$	Normalizing constant
N	Number of nodes in the networks (Network size)
z	Size of the network(side of a square operational area)
λ_n	Mean traffic of the network
b_s	Indicating backoff stages(the number of tries before the successful transmission of the packet)

Appendix 2: Deriving second moment of $E[T_w{}^2]$

In Sect. 3.1.2, the second moment of T_w can be further simplified. The detailed computations are depicted here.

$$E[T_w^2] = \frac{d^2 G_{b_s}(z)}{dz^2}\Big|_{z=1} + E(T_w)$$

$$\begin{aligned} E[T_w^2] =\ & T_s^2(1-c)q(1-p_{tr}(n-1)) \\ & + \sum_{i=1}^{m}\left(T_s^2 + T_c^2 i^2 + \frac{\bar{\sigma}^2}{4}\sum_{j=0}^{i-1}(W_j-1)^2 + 2T_c T_s i \right. \\ & \left. + 2T_s\frac{\bar{\sigma}}{2}\sum_{j=1}^{i-1}(W_j-1) + iT_c\bar{\sigma}\sum_{j=0}^{i-1}(W_j-1)\right)c^i(1-c) \\ & + (m+1)^2 T_c^2 c^{m+1} + c^{m+1}\frac{\bar{\sigma}^2}{4}\sum_{j=0}^{m-1}(W_j-1)^2 \\ & + c^{m+1}(m+1)T_c\bar{\sigma}\sum_{j=0}^{m-1}W_j - 1 \end{aligned}$$

$$\begin{aligned} E[T_w^2] =\ & T_s^2(1-c)q(1-p_{tr}(n-1)) + T_s^2(c - c^{m+1}) \\ & + 2T_s T_c\left(-mc^{m+1} - \frac{c(1-c^m)}{1-c}\right) \\ & + T_c^2\left(\frac{2mc^{m+1}}{1-c} - m^2 c^{m+1} + \frac{c(c+1)(1-c^m)}{(1-c)^2}\right) \\ & + \frac{\bar{\sigma}^2}{4}\left(c\sum_{j=0}^{m-1}(W_j-1)^2 c^j - c^{m+1}\sum_{j=0}^{m-1}(W_j-1)^2\right) \\ & + T_s\bar{\sigma}\left(c\sum_{j=0}^{m-1}\frac{W_j-1}{2}c^j - c^{m+1}\sum_{j=0}^{m-1}W_j - 1\right) \\ & + \bar{\sigma}T_c\left(\sum_{j=0}^{m-1}(W_j-1)(j+1)c^{j+1}\right. \\ & \left. + \sum_{i=j+2}^{m}\sum_{j=0}^{m-2}(W_j-1)c^j - mc^{m+1}\sum_{j=0}^{m-1}(W_j-1)\right) \\ & + (m+1)^2 T_c^2 c^{m+1} + c^{m+1}\frac{\bar{\sigma}^2}{4}\sum_{j=0}^{m-1}(W_j-1)^2 \\ & + c^{m+1}(m+1)T_c\bar{\sigma}\sum_{j=0}^{m-1}(W_j-1) \end{aligned}$$

$$
\begin{aligned}
E[T_w^2] = \ & T_s^2(1-c)q(1-p_{tr}(n-1)) \\
& + T_s^2(c - c^{m+1}) + 2T_s T_c\left(-mc^{m+1} - \frac{c(1-c^m)}{1-c}\right) \\
& + T_c^2\left(\frac{2mc^{m+1}}{1-c} + (2m+1)c^{m+1} + \frac{c(1+c)(1-c^m)}{(1-c)^2}\right) \\
& + \frac{-\bar{\sigma}}{4}\sum_{j=0}^{m-1}(W_j-1)^2 c^{j+1} + T_s\bar{\sigma}\sum_{j=0}^{m-1}(W_j-1)c^{j+1} \\
& + (T_c - T_s)\bar{\sigma}c^m \sum_{j=0}^{m-1} W_j - 1 \\
& + T_c\bar{\sigma}\sum_{j=0}^{m-1}(W_j-1)(j+1)c^{j+1} \\
& + T_c\bar{\sigma}\sum_{i=j+2}^{m}\sum_{j=0}^{m-2}(W_j-1)c^j
\end{aligned}
$$

We derive the average distance between two nodes inside a cell with radius r_n. If r represents the distance from the center of the cell, then the average distance using infinitesimal rings could be formulated as:

$$
\bar{r}_{i,j} = \frac{\int_0^{r_n} 2\pi r(r dr)}{\pi r_n^2} = \frac{2}{3}r_n.
$$

References

1. Takano, Y., & Liou, K.-N. (1989). Solar radiative transfer in cirrus clouds. Part I: Single—scattering and optical properties of hexagonal ice crystals. *Journal of Atmosphere Science, 46*, 3–19.
2. Ozdemir, M., & McDonald, A. B. (2004). A queuing theoretic model for IEEE 802.11 DCF using RTS/CTS. In *Proceedings of IEEE workshop LANMAN* (pp. 33–38).
3. Shabdiz, F. A., & Subramanian, S. (2006). Analytical models for single-hop and multi-hop ad hoc networks. *Mobile Networks and Applications, 11*, 75–90.
4. Marsic, I. (2006). *Wireless networks: Local and ad hoc networks*. New Jersey: Rutgers University Press.
5. Li, J., Blake, C., Couto, D. S. D., Lee, H. I., & Morris, R. (2001). Capacity of ad hoc wireless networks. *In Proceedings of the 7th annual international conference on mobile computing and networking* (pp. 61–69). New York, NY, USA.
6. Bianchi, G. (2000). Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas of Communications, 18*(3), 535–547.
7. Wu, H., Peng, Y., Long, K., Cheng, S., & Ma, J. (2002). Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement. In *In Proceedings of IEEE INFOCOM* (pp. 599–607).
8. Ergen, M., & Varajya, P. (2005). Throughput analysis and admission control for IEEE 802.11a. *ACM MONET, 10*(5), 705–716.
9. Tickoo, O., & Sikdar, B. (2004). Queuing analysis and delay mitigation in IEEE 802.11. In *Proceedings of IEEE INFOCOM, Hong Kong* (pp. 1404–1413).
10. Bisnik, N., & Abouzeid, A. (2009). Queuing network models for delay analysis of multihop wireless ad hoc networks. *Journal of Ad hoc Networks, Elsevier, 7*(1), 79–97.
11. Tickoo, O., & Sikdar, B. (2008). Modeling queuing and channel access delay in unsaturated IEEE 802.11 random access MAC based wireless networks. *IEEE/ACM Transactions on Networking, 16*(4), 878–891.
12. Vassis, D., & Kormentzas, G. (2008). Performance analysis of IEEE 802.11 ad hoc networks in the presence of exposed terminals. *Journal of Ad Hoc Networks, Elsevier, 6*, 474–482.
13. Tobagi, F., & Kleinrock, L. (1977). Packet switching in radio channels: Part IV—stability considerations and dynamic control in carrier sense multiple access. *IEEE Transaction of Communication, 25*(10), 1103–1119.
14. Kleinrock, L. (1976). *Queuing systems, Vol. 2: Computer applications*. New York: Wiley.
15. Bertsekas, D., & Gallager, R. (1992). *Data networks, 2nd edn*. Upper Saddle River: Prentice-Hall International Editions.
16. Reed, D. A., & Fujitomo, R. M. (1987). *Multicomputer networks: Message based parallel processing*. Cambridge: MIT Press.
17. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification (1999). *IEEE Std. 802.11*, 1999.
18. Zhai, H., Kwon, Y., & Fang, Y. (2004). Performance analysis of IEEE 802.11 MAC protocol in wireless LAN. *Wiley Journal of Wireless Communication and Mobile Computing, 4*(8), 917–931.
19. Bisnik, N., & Abouzeid, A. (2006). Delay and throughput in random access wireless mesh networks. In *Proceedings of 2006 IEEE international conference on communications (ICC 2006)*.
20. Zheng, Y., Lu, K., Wu, D., & Fang, Y. (2006). Performance analysis of IEEE 802.11 DCF in imperfect channels. *IEEE Transactions on Vehicular Technology, 55*(5).
21. Kumar, A., Manjunath, D., & Kuri, J. (2004). *Communication networking: An analytical approach, chapter 8* (pp. 456–476). Morgan Kaufman Publishers.
22. Pythagor simulation tool. Available on line at URL: http://www.icsd.aegean.gr/telecom/pythagor/index.htm.

Author Biographies

E. Ghadimi received his B.Sc. degree in Electrical and Computer Engineering from University of Tehran University, Tehran, Iran, in 2005, and his M.Sc. degree in Computer Engineering from the same university, in 2007. From graduation till 2009, he was with the IPM school of Computer Science, Tehran, Iran, as a research student. He is now a Ph.D. candidate at the Department of Computing Science, KTH, Sweden. His research interests include high-performance computer architecture, parallel computing, interconnection networks, and performance modelling/ evaluation.

A. Khonsari received the B.Sc. degree in electrical and computer engineering from Shahid-Beheshti University, Iran, in 1991, and M.Sc. degree in computer engineering from the Iran University of Science and Technology, Iran, in 1996 and Ph.D. degree in computer science from the University of Glasgow, UK, in 2003. He is currently an assistant professor in the Department of Electrical and Computer Engineering, University of Tehran, Iran and a researcher in School of Computer Science, Institute for Studies in Theoretical Physics and Mathematics (IPM), Iran. His research interests are performance modelling/evaluation, mobile and ubiquitous computing, communication networks and distributed systems, and high performance computer architecture.

A. Diyanat received his B.Sc. degree in Electrical and Computer Engineering from University of Tehran University, Tehran, Iran, in 2009. He is now a M.Sc. student at Sharif University of technology, Tehran, Iran. His main research interests include telecommunication systems, distributed optimization, scalable multimedia coding and analytical modelling of wireless ad hoc networks.

M. Farmani received his B.Sc. degree in Electrical and Computer Engineering from University of Tehran, Tehran, Iran, in 2009 and is currently pursuing his M.Sc. degree at the same university. His research interests include distributed optimization and performance evaluation of wireless networks and multimedia coding.

N. Yazdani got his B.S. degree in Computer Engineering from Sharif University of Technology, Tehran, Iran. He worked in Iran Telecommunication Research Center (ITRC) as a researcher and developer for few years. To pursue his education, he entered to Case-Western Reserve Univ., Cleveland, Ohio, USA, later and graduated as a Ph.D. in Computer Science and Engineering. Then, he worked in different companies and research institutes in USA. He joined the ECE Dept. of Univ. of Tehran, Tehran, Iran, as an Assistant Professor in September 2000. His research interest includes Networking, packet switching, access methods, Operating Systems and Database Systems.