Performance analysis of IEEE 802.11 DCF with stochastic reward nets

R. Jayaparvathy¹, S. Anand², S. Dharmaraja³,*⁺ and S. Srikanth¹

¹AU-KBC Research Center, Chennai, India
²Samsung India Software Operations, Bangalore, India
³Department of Mathematics, Indian Institute of Technology, New Delhi, India

SUMMARY

In this paper, we present a performance study to evaluate the mean delay and the average system throughput of IEEE 802.11-based wireless local area networks (WLANs). We consider the distributed coordination function (DCF) mode of medium access control (MAC). Stochastic reward nets (SRNs) are used as a modelling formalism as it readily captures the synchronization between events in the DCF mode of access. We present a SRN-based analytical model to evaluate the mean delay and the average system throughput of the IEEE 802.11 DCF by considering an on–off traffic model and taking into account the freezing of the back-off counter due to channel capture by other stations. We also compute the mean delay suffered by a packet in the system using the SRN formulation and by modelling each station as an M/G/1 queue. We validate our analytical model by comparison with simulations. Copyright © 2006 John Wiley & Sons, Ltd.

Received 16 September 2005; Revised 6 April 2006; Accepted 27 April 2006

KEY WORDS: WLANs; IEEE 802.11; DCF; stochastic reward net; mean delay; throughput

1. INTRODUCTION

The IEEE 802.11 standard for wireless local area networks (WLANs) has become extremely popular due to high data rates and flexible wireless access [1]. The standard defines two modes of operation; the distributed co-ordination function (DCF) and the point co-
ordination function (PCF) [2]. The DCF is a distributed medium access control (MAC) protocol, which is suitable for ad hoc mode of operation, whereas, PCF is a centralized MAC protocol. Several simulation and analytical studies have been made to evaluate the IEEE 802.11 DCF performance [3–12].

Chhaya and Gupta [3] presented a renewal model to study the effect of hidden nodes in IEEE 802.11 with power control. Bianchi [4] proposed an analytical model to obtain the saturation throughput of the IEEE 802.11 DCF. An embedded Markov chain-based approach was presented, considering binary exponential back off (BEB). In References [5, 6], the analytical model in Reference [4] was modified adapting it to back-off variants. Chatzimisios et al. [7] presented an analytical model to evaluate the saturation throughput and the frame delay of the IEEE 802.11 DCF under traffic conditions that correspond to the maximum load. Cali et al. [8] considered a geometrically distributed back-off and derived an analytical formula for the MAC protocol capacity. Carvalho and Aceves [9] extended the model in References [4, 8] to present a model for the mean and the variance of the time a node spends in back off and freezing. As in References [4, 8], the authors considered an always on model and saturated traffic conditions. Foh et al. [10] provided a queuing model to evaluate the DCF performance. The model took into account, the statistical characteristics of the protocol operations. Bianchi and Tinnirello [11] presented an approach based on elementary conditional probability arguments to evaluate the throughput and delay performance of the IEEE 802.11 DCF. Wu et al. [12] enhanced the analysis in Reference [4] to obtain the goodput\(^\text{\textsuperscript{1}}\) of the system. Developing an analytical model capturing all the DCF protocol operations is challenging, because of the inherent synchronization required between the events. The advances in the theory of stochastic petri nets (SPN) [13] motivated SPN-based analysis of the IEEE 802.11 DCF. German and Heindi [14] formulated a detailed SPN model for performance evaluation of the IEEE 802.11-based WLAN. The authors derived two compact and analytically tractable models. Although the SPN model in Reference [14] captures most of the relevant system aspects, it does not exactly capture the freezing of the back-off counter at a station when some other station captures the channel.

In this paper, we propose a stochastic reward net (SRN)-based approach to model the IEEE 802.11 DCF. Our model captures the freezing of the back-off counter in addition to capturing the other system aspects as in Reference [14]. The always-on-arrival model that has been considered in References [4, 8, 9] has been modified in our model to be an on–off model. The SRN formulation not only provides the probability of retransmission of a packet due to collisions (and hence, the average system throughput) [15], but also provides the mean delay suffered by the first packet (i.e. the packet at the head of line (HOL)) at every station. To compute the mean delay of the subsequent packets at a station, we model each station as an \(M/G/1\) queue [16], with the mean service time to be the mean delay suffered by the HOL packet. We validate our analytical model by comparison with simulations.

The rest of the paper is organized as follows. In Section 2, we explain the IEEE 802.11 DCF operation. In Section 3, we present the system model. In Section 4, we develop and explain the SRN model for the underlying system, and the performance analysis. Section 5 provides the numerical results and Section 6 provides the conclusions.

\(^{1}\)Goodput has been defined in the reference paper [12]. Goodput refers to the bandwidth the user actually receives.
2. IEEE 802.11 DCF

The IEEE 802.11 PCF and DCF MAC protocols are slotted protocols. In the IEEE 802.11 DCF, the multiple access mechanism is carrier sense multiple access with collision avoidance (CSMA/CA). Carrier sensing is performed both at the physical layer (referred to as physical carrier sensing) and at the MAC layer (also known as virtual carrier sensing). A handshaking mechanism with request-to-send (RTS) and clear-to-send (CTS) signals is used to reserve the channel before data transmission. This is explained as follows.

Consider a network as shown in Figure 1 in which, transmissions from station $S$ can be received by stations $A$, $B$, $C$, $D$ and $X$. Similarly, transmission from station $D$ can be received by stations $S$, $X$ and $Y$. If $S$ has data to be transmitted to $D$, station $S$ senses the channel. If the channel is sensed idle for a time interval equal to a DCF inter-frame spacing (DIFS) [1], then it sends an RTS signal, which specifies the destination station, $D$, and the size of the data packet, which, in turn, specifies the time up to which the channel will be busy. The RTS is received not only by station $D$, but also by stations $A$, $B$, $C$ and $X$. On receiving the RTS, stations $A$, $B$, $C$ and $X$ set their network allocation vectors (NAV) based on the packet size. Station $D$ responds with a CTS, which is heard by station $S$ and also heard by stations $X$ and $Y$. The CTS piggybacks the packet size information. Station $X$ ignores the CTS, while station $Y$ sets its NAV based on the packet size information in the CTS. The NAV at each station gives the minimum amount of time the station needs to wait before it begins to sense the channel. Once the CTS is received by station $S$, data packets (DATA) are transmitted from station $S$ to station $D$ and station $D$ responds with an acknowledgment (ACK). The handshake mechanism using RTS, CTS, DATA and ACK is called as the four-way handshake, and is shown to overcome the hidden node problem [2]. The DCF access mechanism is depicted in Figure 2.

It is observed that the four-way handshake mechanism described above, could result in collision of RTS packets [2]. Binary exponential back-off (BEB) is used to reduce collisions [1]. The BEB procedure works as follows: When station $S$ has data packets to be transmitted to

![Figure 1. CSMA/CA in the IEEE 802.11 DCF.](image-url)
station $D$, station $S$ senses the channel. If the channel is sensed idle for a DIFS time interval, then station $S$ generates a random integer, $b$, uniformly distributed in $[0, CW_{\text{min}}]$. A back-off counter is initialized to the random integer, $b$, and is decremented at discrete time slots at the rate of one per slot. The station $S$ transmits RTS only when the back-off counter reaches zero. If the channel is sensed busy when the back-off counter reaches a value $b' > b$ (i.e. if the channel is sensed busy when the back-off counter is being decremented but before it reaches zero), then the back-off counter is frozen at $b'$, and the station continues to sense the channel. The back-off counter is then decremented (starting from $b'$) only if the channel is sensed idle for a DIFS period. Collision of packets occur when two or more stations generate random integers such that the respective back-off counters reach zero at the same slot. If after transmission (i.e. after the back-off counter reaches zero) the station $S$ does not receive the CTS within a time period specified by a short inter-frame spacing (SIFS) + CTS time [1], then the RTS is assumed to have collided with the transmission from another station and station $S$ is not allowed to transmit. After each unsuccessful transmission the back-off counter is initialized to a new random integer $b_2$, which is uniformly distributed in $[0, 2CW_{\text{min}}]$. In general, the back-off counter is initialized to a random integer which is uniformly distributed in $[0, CW_k]$, where $k > 0$ is called as the back-off stage or the retransmission number. The value $k = 0$ represents the first time transmission and value $k > 1$ represents the $k$th retransmission. Also, $CW_k$ is given by $CW_k = \min(2^k(CW_{\text{min}} + 1) - 1, CW_{\text{max}})$, where $CW_{\text{min}}$ and $CW_{\text{max}}$ are specified by the IEEE 802.11 standard [1]. The above procedure (i.e. channel-sensing and back-off) is continued till the CTS is received successfully. If the CTS is not received successfully in 7 transmission attempts of the RTS, then the packet is dropped from the system. Upon successful reception of the CTS, the packet is transmitted to station $D$, and if the ACK is not received by station $S$ within an SIFS time interval + ACK time, then the packet is assumed to be corrupted due to physical layer impairments. Corrupted packets are retransmitted following the four-way handshake mechanism with BEB. If the ACK is not
received successfully for 7 transmission attempts of the packet, then the packet is dropped from the system. It is also observed that the BEB only reduces the probability of collision, but does not eliminate collisions, i.e. the probability of collision in a system with BEB is also non-zero.

3. SYSTEM MODEL

In a typical WLAN, stations are distributed in a grid of dimension $X \times X$. Stations can send and receive data from its one-hop neighbours. If stations have data to transmit to stations other than their one-hop neighbours, then the data packets are routed through the one-hop neighbours. The WLAN network can be modelled as a graph as shown in Figure 3, in which each station is represented as a vertex and two vertices are joined by an edge if the corresponding stations are one-hop neighbours.

We consider a WLAN with $N$ stations. The objective is to obtain an analytical model to evaluate the mean delay and the average system throughput in such a network which deploys the IEEE 802.11 standard and operates in the DCF mode. It is desired to obtain an analytical model, which takes into account the dropping of packets that exceed the maximum number of retrials, and also takes into account the freezing of the back-off counter when the channel is sensed busy.

Calls or data packets arrive at each station at random epochs of time. When a data packet arrives at a station, the packet is transmitted following the CSMA/CA with BEB as explained in Section 2. Each station maintains a buffer. Data packets arriving at a station are buffered till they get access to the channel. Newly arriving data packets at a station are added to the buffer on a first come first serve (FCFS) basis, whereas, a packet that is retransmitted due to collision is treated as the first packet or the HOL packet of the buffer. The delay of a packet is defined as the

![Figure 3. A typical wireless LAN with 20 stations.](image)
time spent by a packet in the system till it is successfully transmitted (i.e. till the packet is received correctly by the destination station, and the corresponding ACK is received correctly by the source station). The delay includes the delay suffered by all retransmissions of the packet. The mean delay of a packet is defined as the average delay suffered by a packet. The average system throughput is defined as the fraction of time successful transmissions take place in the channel, i.e. the fraction of time the channel is busy and collision free.

It is observed that the mean delay is the sum of the mean queuing delay at the buffer of each station, the delay due to the DIFS, SIFS, RTS, CTS and ACK, and the delay due to back off and freezing of the back-off counter when the channel is captured by another station. Hence, it is necessary to obtain a model that captures the back off and the freezing of the back-off counter, which is our focus in this paper.

We make the following assumptions to carry out the performance study.

- There are \( N = 10 \) stations in the system, distributed in a square grid of dimension \( X \times X \).
- The call arrival process at any station is a Poisson process with mean arrival rate, \( \lambda \). (Since we are considering the basic IEEE 802.11 DCF, we consider the best-effort-data-only traffic. Data bursts consist of active and idle periods. In practice a data burst is a data packet of variable length, for example an IP packet with zero idle time between a finite set of consecutive packets. A call is a data burst whose arrival process is Poisson with mean arrival rate \( \lambda \).)
- The call holding times are exponentially distributed with mean \( 1/\mu \) seconds.
- All the stations are statistically independent and behave identically.
- The errors due to multipath fading are neglected. \(^5\)
- The wireless stations are assumed to have very low mobility, or no mobility.

### 4. PERFORMANCE EVALUATION

In this section, we present the SRN model to obtain the mean delay and the average system throughput. We proceed as follows: In Section 4.1, we present the SRN representation to model the DCF operation. In Section 4.2, we present the performance analysis to obtain the mean delay and the average system throughput from the SRN model. An introduction to SRN and its components is provided in Appendix A.

#### 4.1. Stochastic reward net model

In this subsection, we develop and explain the SRN model for the IEEE 802.11 DCF. We present an SRN model based on the decomposition approach [17], which is explained as follows. Since all the stations are independent and behave identically, the characteristics of the system can be captured by studying the behaviour of the reference station in detail, and the cumulative behaviour of all the other stations. Thus, the SRN model representing the IEEE 802.11 DCF is as shown in Figure 4.

---

\(^5\)Our model can be easily extended to incorporate fading effects.
It is observed from Figure 4, that the model consists of two parts, A and B. Part A, represents the events at the reference station, and part B, represents the events at all the other stations. The places, transitions, and the guard functions associated with the SRN are listed in Tables I–III.

![Stochastic reward net model for the IEEE 802.11 DCF.](image)

**Table I. List of places.**

<table>
<thead>
<tr>
<th>Reference station</th>
<th>Place</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{rs}$</td>
<td>Packets that arrive at the buffer</td>
</tr>
<tr>
<td></td>
<td>$P_{DIFS}$</td>
<td>Packet waiting for DIFS duration</td>
</tr>
<tr>
<td></td>
<td>$P_{BS}$</td>
<td>Packets in the back-off state</td>
</tr>
<tr>
<td></td>
<td>$P_{channel}$</td>
<td>State of the channel</td>
</tr>
<tr>
<td></td>
<td>$P_{trans}$</td>
<td>Packet ready for first time transmission</td>
</tr>
<tr>
<td></td>
<td>$P_{retry}$</td>
<td>Packet waiting for retransmission</td>
</tr>
<tr>
<td></td>
<td>$P_{os_send}$</td>
<td>Packet ready to be transmitted on the channel</td>
</tr>
<tr>
<td></td>
<td>$P_{rand1}$</td>
<td>Number of tokens for random back off</td>
</tr>
<tr>
<td></td>
<td>$P_{rand2}$</td>
<td>To generate random number of tokens in $P_{rand1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other stations</th>
<th>Place</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{os}$</td>
<td>Packets that arrive at the buffer</td>
</tr>
<tr>
<td></td>
<td>$P_{os_send}$</td>
<td>Packet ready to be transmitted on the channel</td>
</tr>
</tbody>
</table>

It is observed from Figure 4, that the model consists of two parts, $A$ and $B$. Part $A$, represents the events at the reference station, and part $B$, represents the events at all the other stations. The places, transitions, and the guard functions associated with the SRN are listed in Tables I–III.
It is observed that the intervals between events in the IEEE 802.11 DCF MAC protocol could be either deterministic (e.g. DIFS, SIFS, RTS, CTS, ACK, etc.), or random (e.g. back-off time, inter-arrival time, etc.). However, the SRN assumes exponentially distributed firing times for all the timed transitions. We use timed transitions to represent not only the exponentially distributed random intervals, but also to represent all the deterministic intervals, and the random intervals that are not exponentially distributed. This is done to enable the modelling using SRN, and validated in Section 5 by comparison with simulations.

### 4.1.1. Places and transitions of the SRN

- Transition $T_{rs}$ represents the arrival of packets at the reference station. When transition $T_{rs}$ fires, one token is deposited in place $P_{rs}$. The mean firing time of $T_{rs}$ is the mean inter-arrival time of packets at the reference station (i.e. $\lambda^{-1}$).
- Transition $T_{os}$ represents the arrival of packets at all the other stations. When transition $T_{os}$ fires, one token is deposited in place $P_{os}$. $T_{os}$ has a mean firing time $[(N-1)\lambda]^{-1}$.
- Place $P_{rs}$ represents the buffer of the reference station.
- The cumulative buffer of all the other stations is represented by the place $P_{os}$.

It is observed that the intervals between events in the IEEE 802.11 DCF MAC protocol could be either deterministic (e.g. DIFS, SIFS, RTS, CTS, ACK, etc.), or random (e.g. back-off time, inter-arrival time, etc.). However, the SRN assumes exponentially distributed firing times for all the timed transitions. We use timed transitions to represent not only the exponentially distributed random intervals, but also to represent all the deterministic intervals, and the random intervals that are not exponentially distributed. This is done to enable the modelling using SRN, and validated in Section 5 by comparison with simulations.

### 4.1.1. Places and transitions of the SRN

- Transition $T_{rs}$ represents the arrival of packets at the reference station. When transition $T_{rs}$ fires, one token is deposited in place $P_{rs}$. The mean firing time of $T_{rs}$ is the mean inter-arrival time of packets at the reference station (i.e. $\lambda^{-1}$).
- Transition $T_{os}$ represents the arrival of packets at all the other stations. When transition $T_{os}$ fires, one token is deposited in place $P_{os}$. $T_{os}$ has a mean firing time $[(N-1)\lambda]^{-1}$.
- Place $P_{rs}$ represents the buffer of the reference station.
- The cumulative buffer of all the other stations is represented by the place $P_{os}$.

It is observed that the DCF operations apply only to packets that are at the HOL of the buffer at each station. Hence, it is sufficient to model the DCF operations of the HOL packets at all stations. Hence, an inhibitor arc with multiplicity $K_{rs} = 1$ is added between place $P_{rs}$ and transition $T_{rs}$. Similarly, an inhibitor arc with multiplicity $K_{os} = N - 1$ is added between place $P_{os}$ and transition $T_{os}$, where $N$ is the number of stations.

The place $P_{DIFS}$ represents the channel sensing by the reference station. Since each station can transmit at most one packet at a time, a token must be added to place $P_{DIFS}$ only when there is
no other packet at the station that is sensing the channel, i.e. the new packet is the HOL packet at the reference station. This is represented by the immediate transition $t_{\text{DIFS}}$ and the inhibitor arc between $t_{\text{DIFS}}$ and $P_{\text{DIFS}}$. The timed transition $T_{\text{DIFS}}$ represents the channel sensing for a DIFS interval.

The back-off counter initialization is represented by $P_{\text{BS}}$. If the back-off counter is to be initialized to $n$, then $n$ tokens are to be deposited in $P_{\text{BS}}$ when $T_{\text{DIFS}}$ fires. Therefore, the multiplicity of the directed arc $(T_{\text{DIFS}}, P_{\text{BS}})$ should be a random integer $n$ which is uniformly distributed in $[0, CW_k]$, for the $k$th trial. The number of states in the underlying continuous time Markov chain (CTMC) of the SRN is $O(N^2n)$, where $N$ is the number of stations. To make $n$ to be a random value, places $P_{\text{rand1}}, P_{\text{rand2}}$ and transitions $T_{\text{rand1}}$ and $T_{\text{rand2}}$ are incorporated in the SRN. Initially a random number of tokens are deposited in $P_{\text{rand1}}$. The transitions $T_{\text{rand1}}$ and $T_{\text{rand2}}$ are assigned different probabilities of firing to enable the number of tokens in $P_{\text{rand1}}$ to vary. The multiplicity of the arc $(T_{\text{DIFS}}, P_{\text{BS}})$, $n$ is decided by the number of tokens in $P_{\text{rand1}}$ using a function $n = \#P_{\text{rand1}}$.

Each decrement of the back-off counter is represented by the transition $T_{\text{BS}}$. The mean firing time of $T_{\text{BS}}$ is $X_{\text{slot}}$, where $X_{\text{slot}}$ is the slot time as defined by the IEEE 802.11 standard [1]. $P_{\text{trans}}$ represents a packet in the back-off state. The inhibitor arc between $T_{\text{BS}}$ and $P_{\text{trans}}$ ensures that the back-off counter does not decrement below zero.

---

Figure 5. Mean delay: bit rate = 2 Mbps, packet size = 512 bytes.
The place $P_{\text{channel}}$ represents the state of the channel, such that, there is one token in $P_{\text{channel}}$ if the channel is idle, and there are no tokens in $P_{\text{channel}}$ if the channel is busy. The transition $t_{\text{transF}}$ indicates that a packet is ready for its first time transmission after the back-off counter reaches zero. The arc $(P_{\text{trans}}, t_{\text{transF}})$ is assigned a multiplicity $n$ so that a packet is not transmitted until the back-off counter reaches zero. The arc $(P_{\text{channel}}, t_{\text{transF}})$ is added to indicate successful packet transmission only if the channel is idle when the back-off counter reaches zero.

The place $P_{\text{rs\_send}}$ represents transmission of the packet with mean holding time given by the mean firing time of transition $T_{\text{rs\_send}}$ (i.e. $T_{\text{rs\_send}}$ is assigned a mean firing time, $\mu^{-1}$). The cumulative effect of similar operations at all the other stations is represented by $T_{\text{CT}}, P_{\text{os\_send}}$ and $T_{\text{os\_send}}$.

Packets undergo collision (and hence, retransmission) when the back-off counter of the reference station and that of at least one other station reach zero simultaneously. The retransmissions of packets at the reference station are represented by the place, $P_{\text{retry}}$ and the transition, $t_{\text{retry}}$. The firing of $t_{\text{retry}}$ represents a collision, and the inhibitor arc between $P_{\text{channel}}$ and $t_{\text{retry}}$ enables $t_{\text{retry}}$ when a collision occurs. Subsequent retransmission attempts of a packet are represented by the transition $t_{\text{transR}}$.

The number of retransmission attempts of a packet is given by the number of tokens in $P_{\text{retry}}$. As mentioned in Section 2, the retransmitted packets must undergo a similar sequence of operations. However, modelling retransmissions in the SRN results in a complex underlying CTMC, which cannot be solved. Therefore, instead of retransmitting the same packet, we pass the next token in the place $P_{\text{rs}}$ by firing of the transition $t_{\text{DIFS}}$.

Figure 6. Mean delay: bit rate = 11 Mbps, packet size = 512 bytes.
The transition $t_{\text{lost}}$ represents dropping of packets from the system due to exceeding the maximum retransmission limit, which is taken to be 7 for the IEEE 802.11 DCF [1]. The multiplicity of the arc $(P_{\text{retry}}, t_{\text{lost}})$ is hence assigned a value 7.

4.1.2. Guard functions. It is also necessary to model the following operations of the IEEE 802.11 DCF MAC protocol, apart from those explained in Section 4.1.1.

- The channel should be sensed idle for the entire DIFS duration for every packet transmission (i.e. the first time transmission and all retransmissions) before entering the back-off state.
- When the channel is sensed busy during the back-off state, then the back-off counter must be frozen until the channel is again sensed idle for a DIFS duration.

The above-mentioned events are modelled by assigning appropriate guard functions to some of the transitions. The guard functions used in our SRN model are listed in Tables II and III, and explained below.

- Let the number of tokens in any place $P$ in the SRN be denoted by $\#P$. The transition $t_{\text{DIFS}}$ is assigned a guard function $\#P_{\text{BS}} = 0 \& \& \#P_{\text{trans}} = 0$, where $\&\&$ represents logical AND. The assigned guard function allows $t_{\text{DIFS}}$ to fire only if the packet is at the HOL of the buffer of the reference station.

Figure 7. Mean delay: bit rate = 54 Mbps, packet size = 512 bytes.
The guard function for the transition $T_{DIFS}$ is $\#P_{\text{channel}} = 1 \& \& \#P_{\text{trans}} = 0$. This guard function allows $T_{DIFS}$ to fire (i.e. allows a packet to enter the back-off state) only when the channel is idle for a DIFS period, and when the packet is at the HOL of the buffer.

The guard function associated with transition $T_{BS}$ is $\#P_{\text{channel}} = 1$. This function models the freezing of the back-off counter when the channel is sensed busy while in the back-off state.

4.2. Mean delay and average system throughput

The underlying CTMC of the SRN explained in Section 4.1 can be obtained from the extended reachability graph (ERG). ERG is a directed graph which is explained in Appendix A. The average number of tokens in a place $P$ can be determined from the steady state probability of occupancy of each state in the CTMC, which can be obtained by using standard software packages for SRNs like TimeNET [18], or SHARPE [19], or SPNica [20]. The packages also enable calculation of the average throughput of a transition, which is defined as the average rate at which tokens are deposited by the transition in its output places, i.e. if $Y(t)$ is the average number of tokens deposited by the transition $T$ is all its output places up to time $t$, then the throughput of the transition $T$, $\eta_T$ is defined as

$$\eta_T = \lim_{t \to \infty} \frac{Y(t)}{t}$$

![Figure 8](image_url)  
Figure 8. Average system throughput: bit rate = 2 Mbps, packet size = 512 bytes.
For a mean packet arrival rate \( \lambda \), and mean packet holding time \( \mu^{-1} \), the average throughput of a station, \( \eta \), is given by

\[
\eta = \left( \frac{\lambda}{\mu} \right) \frac{\eta_{T_{rs\_send}}}{\eta_{T_{rs}}}
\]

which is also the average system throughput because the behaviour of the stations are independent and identically distributed. The mean delay of the HOL packet of each station, \( \bar{D}_{HOL} \), is the sum of the mean packet holding time and the sum of the mean delays undergone by the HOL packet at every stage of the DCF operation. This is obtained by measuring the mean delay of the HOL packet at places \( P_{rs} \), \( P_{BS} \) and \( P_{rs\_send} \) in the SRN. Let the average number of tokens in place \( P \) be \( \#P \). By associating unit reward to all the markings, \( \bar{D}_{HOL} \) can be obtained by using Little’s theorem [21] as

\[
\bar{D}_{HOL} = \frac{\#P_{rs}}{\eta_{T_{rs}}} + \frac{\#P_{BS}}{\eta_{T_{DIFS}}} + \frac{\#P_{rs\_send}}{\eta_{T_{BS}}} + \frac{1}{\mu}
\]

The rest of the buffer at the reference station is modelled as an \( M/G/1 \) queue with mean service time to be \( \bar{D}_{HOL} \). The mean packet delay, \( \bar{D} \) can then be obtained by applying the

![Figure 9. Average system throughput: bit rate = 11 Mbps, packet size = 512 bytes.](image-url)
Pollackzek-Kinchine mean value formula [21] as

$$\bar{D} = \bar{D}_{\text{HOL}} \left[ 1 + \frac{\rho_b}{2(1 - \rho_b)} (1 + C_D^2) \right]$$

(3)

where $\rho_b \triangleq \lambda \bar{D}_{\text{HOL}}$, and if the delay of the HOL packet is represented by the random variable, $D$, then

$$C_D^2 = \frac{E[D^2]}{\bar{D}_{\text{HOL}}}$$

(4)

To obtain the value of $E[D^2]$, it is necessary to obtain the distribution of $D$. For simplicity of analysis, we neglect the dropping of packets due to exceeding the delay limit specified by the Max Transmit MSDU lifetime. This approximation is valid because, the probability of packet delay exceeding the Max Transmit MSDU lifetime is small for the loads under consideration. $E[D^2]$ can then be obtained as

$$E[D^2] = 2 \left( \frac{\# P_{\text{received}}}{\eta_{\text{ris}}} \right)^2$$

(5)

The mean delay $\bar{D}$ obtained from (2)–(5) is the mean delay suffered by any packet in the system because, all the stations are independent and behave identically.

Figure 10. Average system throughput: bit rate = 54 Mbps, packet size = 512 bytes.
5. RESULTS AND DISCUSSION

In this section we present the numerical results obtained using the SRN model for the IEEE 802.11 DCF with $N = 10$ stations in a square of dimensions $500 \times 500$, such that all stations are adjacent to each other. We present the mean delay and the average system throughput as a function of the virtual traffic load into the system in Erlangs, $V$, which is defined as $V = N\bar\lambda L/B$, where $L$ is the mean packet length, $B$ is the channel bit rate and $\bar\lambda$ represents the mean packet arrival rate. We use the SHARPE package [19] for specifying the SRN, and solving the underlying CTMC. We compare the analytical results with those obtained by simulation. We consider the IEEE 802.11a standard [2] operating at data rates of $B_1 = 2$, $B_2 = 11$ and $B_3 = 54$ Mbps. The arrival rate, $\bar\lambda$, is varied to obtain values of $V$ in the range 0.1 to 7 Erlangs. We use the following values for the system parameters in our computations.

- Packet size, $L = 512$ bytes
- RTS = 20 bytes
- CTS = 14 bytes
- ACK = 14 bytes
- SIFS = 16 $\mu$s

![Figure 11. Mean delay vs throughput: bit rate = 2 Mbps, packet size = 512 bytes.](image-url)
DIFS = 34 µs
EIFS = 330 µs
\( CW_{\text{min}} = 15 \)
\( CW_{\text{max}} = 1023 \)
slot-time, \( X_{\text{slot}} = 9 \) µs

Figures 5–7 present the mean packet delay in a system with data rate of 2, 11 and 54 Mbps, respectively. It is observed that the analytical results match closely with the simulation results, thus validating the SRN model and the approximations made in the analysis. It is observed that for a same load, the mean packet delay decreases as the channel bit rate increases, e.g. for \( V = 0.8 \) Erlangs, the mean delay is 2.5 ms, 560 µs and 227 µs for bit rates of 2, 11 and 54 Mbps, respectively. This is because, for a given load, the mean delay in a queue decreases with decreasing service time [21].

Figures 8–10 present the average system throughput for bit rates of 2, 11 and 54 Mbps, respectively. It is observed that the throughput is approximately a linear function of the virtual load for small values of \( V \). This is because, for a virtual load \( V \), the throughput is given by \( \eta = V(1 - p_c) \), where \( p_c \) is the probability of collision. However, for small values of \( p_c (p_c \ll 1) \), throughput, \( \eta \approx V \), and hence, a linear function of \( V \). For higher values of \( p_c \), the throughput may not be linear as \( p_c \) is comparable to 1. Since the system behaves like a single server queue, it
is necessary to have $V < N(1 - p_c)$ for stability. We observed instability to occur at values of $V > 4$ for the 11 Mbps system, and $V \geq 1$ for the 54 Mbps system. This is because, as the channel bit rate increases, the mean packet holding time decreases, and hence, the arrival rate increases for the same value of $V$. Therefore, the network reaches closer to the saturation condition (i.e. all stations having packets to transmit). Hence, the probability of the back-off counter at two stations reaching zero simultaneously, increases, thus resulting in more collisions.

Figures 11–13 present the variation of the mean delay with the average system throughput for bit rates of 2, 11 and 54 Mbps, respectively. It can be observed from these curves that, an increase in the mean delay reduces the average system throughput for the values of load mentioned above.

Figure 14 presents the mean packet delay, for various packet sizes for a mean packet arrival rate of 244 packets per second, for a system with data rates of 2, 11 and 54 Mbps. The above results are obtained from our analytical model. From Figure 14, it is observed that with a mean packet arrival rate of 244 packets per second, the 54 Mbps data rate system can support a packet size of 20 KB with a mean packet delay of 6.5 ms, whereas smaller packet sizes of 4 and 0.8 KB are supported by the 11 and 2 Mbps data rate systems, respectively, for the same mean packet delay.

Our model can be modified to account for the Max Transmit MSDU lifetime by adding a new transition similar to $t_{\text{lost}}$, and assigning appropriate guard functions. The SRN representation we presented in this paper, can also be extended to model the IEEE 802.11e enhanced DCF.
(EDCF) [22] by representing multiple classes of traffic using coloured tokens, and assigning appropriate mean firing times and priorities to transitions. Our analysis can be extended to take into account physical layer impairments like bit errors due to fading, by modelling the physical layer to be independent of the MAC, and hence, for a packet error rate $p_e$, the average system throughput is degraded by a factor $(1 - p_e)$, and the mean delay increases by a factor $(1 - p_e)^{-1}$. Mobility of stations can be accounted for, by varying (a) the value of $K_{os}$, (b) the firing time associated with $T_{os}$, and (c) associating appropriate reward functions to the markings.

6. CONCLUSION

We presented an analysis using SRN to evaluate the mean delay and the average system throughput of the IEEE 802.11 DCF MAC protocol. Our model took into account, the effect of freezing of the back-off counter, and the maximum number of packet retrials. We validated our analysis by comparison with simulations. Our model can be extended to analyse multi-hop systems, and systems with multiple classes of traffic. It is also possible to easily apply our model for any generalized arrival and service processes by deploying other packages for the SRN. Our model can also be modified to perform a call admission control based on the service requirement and the channel conditions. The extension of our model using different reward rates and using
coloured petri nets for different classes of traffic is of significant interest and a topic for further investigation.

APPENDIX A: INTRODUCTION TO SRN AND ITS COMPONENTS

Stochastic petri net (SPN) is an extension of petri net (PN) [13], which is a high-level description language for formally specifying complex systems. SPN has been used as a powerful modelling tool in performance, availability and reliability analysis in communication systems [23]. The general components of an SPN are shown in Figure A1.

Definition A.1
A petri net (PN) is a bi-partite directed graph with two types of nodes called places and transitions, that are connected by directed edges or directed arcs. Arcs exist only between places and transitions, i.e. there is no arc between two places or two transitions.

Each place may contain an arbitrary number of tokens. The number of tokens in a place is a non-negative integer. For a pictorial presentation, places are depicted as circles, transitions are represented by bars and tokens are represented by dots or integers in the places. Each directed arc in the bi-partite graph is assigned a weight or a multiplicity, which is a natural number. If the multiplicity of an arc is not specified, then it is taken to be unity.

Definition A.2
For any transition in the PN, arcs directed out of the transition are called as output arcs of the transition. The corresponding places are called as output places of the transition.

Definition A.3
For any transition in the PN, arcs directed into the transition are called input arcs of the transition. The corresponding places are called as input places of the transition.

![Figure A1. Components of an SRN.](image-url)
A transition in a PN is said to be enabled if all of its input places have at least as many tokens as the multiplicities of the corresponding input arcs.

An inhibitor arc is an undirected arc (between a place and a transition) with multiplicity $k \geq 1$. If the multiplicity of the inhibitor arc is not specified, then it is taken to be unity. An inhibitor arc is pictorially represented as shown in Figure A1.

If there exists an inhibitor arc with multiplicity $k$ between a place and a transition, and if the place has $k$ or more tokens, then the transition is inhibited even if it is enabled. When enabled (and not inhibited), a transition can fire. When a transition fires, it removes from each input place, the number of tokens given by multiplicities of the corresponding input arcs, and adds to each output place, the number of tokens given by the multiplicities of the corresponding output arcs.

Consider a PN containing $M$ places and $N$ transitions.

A marking, $\mathcal{M}(t)$ of a PN is an $M$ tuple, $[m_1(t) \ m_2(t) \ m_3(t) \ \ldots \ m_M(t)]$ of non-negative integers, where $m_i(t)$ denotes the number of tokens in place $i$ ($1 \leq i \leq M$) at any given instant of time, $t$. For a given reference time instant, $t_0$, $\mathcal{M}(t_0)$ is called as the initial marking of the PN.

With respect to a given initial marking, the reachability set is the set of all markings that can be obtained from the initial marking through any possible firing sequences of transitions.

If with each transition, a parameter, $T$, is associated such that, a transition that is enabled at time $t$ cannot fire till time $t + T$, then the parameter $T$ is called as the firing time of the transition.

Stochastic petri nets (SPNs) [24] is an extension of the PNs in which each transition is assigned a firing time. The firing time could be zero or could be a random variable which is exponentially distributed. Transitions with exponentially distributed firing times (drawn as rectangular boxes) are called timed transitions, while transitions with zero firing times (drawn as thin black bars) are called immediate transitions. Immediate transitions have higher priority over timed transitions.

A marking $\mathcal{M}(t)$ in an SPN is called vanishing if at least one immediate transition is enabled at time $t$. A marking which is not vanishing is called a tangible marking.

Let $\mathcal{N}(t)$ be the set of all possible markings at time $t$.

A guard function $g_T(t)$ associated with transition $T$, is a Boolean function defined over $\mathcal{N}(t)$ (i.e. $g_T(t) : \mathcal{N}(t) \rightarrow \{0, 1\}$), such that the transition $T$ does not fire if $g_T(t) = 0$ (even if enabled and not inhibited), and fires if it is enabled, not inhibited and $g_T(t) = 1$. 

Definition A.11
A reward is a non-negative weight associated with each marking. The reward rate of a place $P$ is the weighted average of the number of tokens in $P$ over all markings.

Definition A.12
A stochastic reward net (SRN) is an extension of an SPN that allows extensive marking dependency and expresses complex enabling or disabling conditions for transitions (a) through guard functions, or (b) by assigning one or more reward rate(s) to each tangible marking.

Definition A.13
For a given SPN or SRN, an extended reachability graph (ERG) is a directed graph with the markings of the reachability set as the nodes and a directed edge from node $\mathcal{M}_1(t)$ to $\mathcal{M}_2(t + T)$ if $\mathcal{M}_2(t + T)$ can be obtained from $\mathcal{M}_1(t)$ by firing of a single transition.

To each edge in the ERG, some stochastic information is attached. The stochastic information could be the multiplicative inverse of the mean firing time of the transition, or could be the probability of firing of the transition. If the expected number of transitions that fire in finite time is finite, then it can be shown that a given ERG can be reduced to a homogeneous CTMC [24]. The number of states in the CTMC depends on the number of places, the number of transitions, the set of arcs (including the directed arcs as well as the inhibitor arcs), the multiplicities of the arcs and the initial marking. Therefore, it is necessary to have minimum redundancies in the SRN to reduce the underlying state space, and hence, the computational complexity. Depending on the nature of the system, different approaches like state truncation [25], hierarchical approach [26] and decomposition approach [17], can be applied to reduce the underlying state space for the SRN.

Parameters such as the firing rate of the timed transitions, the multiplicities of input and output arcs, and the reward rate in a marking can be specified as functions of the number of tokens in any place in the SRN. For an SRN, all the output measures are expressed in terms of the expected values of the reward rate functions. To get the throughput, delay or any other performance measure, appropriate reward rates are assigned to the markings of the SRN. As SRN is automatically transformed into a Markov reward model (MRM), and the required measures of the SRN can be obtained by either a steady state analysis or a transient analysis of the underlying MRM.

APPENDIX B: NUMBER OF STATES IN THE UNDERLYING CTMC OF THE SRN IN FIGURE 4

Consider the SRN shown in Figure 4. To compute the number of states in the underlying CTMC of the SRN, it is necessary to obtain the underlying CTMC. However, the underlying CTMC is expected to have nine dimensions corresponding to the nine places in the SRN. We make an approximate estimate of the number of states, as follows. Places $P_{rand1}$ and $P_{rand1}$ are not taken into account to compute the CTMC states as they are just used to deposit the $n$ tokens in place $P_{BS}$.

1. Let $m_i$ represent the number of tokens in place $P_i$ at steady state. It is noted that $m_{rs} \leq K_{rs}$ and $m_{DIFS}$ can be either 0 or 1. However, $m_{rs} = 0$ if $m_{DIFS} = 0$, due to the immediate transition, $t_{DIFS}$. This leads to $K_{rs} + 1$ possible states with varying $m_{rs}$ and $m_{DIFS}$.
2. When $T_{\text{DIFS}}$ fires, $n$ tokens are deposited in place, $P_{\text{BS}}$, and this is decremented in steps of 1, following each firing of $T_{\text{BS}}$. Therefore, if $m_{\text{BS}} = k$, then $m_{\text{trans}} = n - k$, for $0 \leq k \leq n$. It is noted that the inhibitor arc between $T_{\text{DIFS}}$ and $P_{\text{BS}}$ prevents states such that $m_{\text{BS}} + m_{\text{trans}} \neq n$. Therefore, there are $n + 1$ possible states for different values of $m_{\text{BS}}$ and $m_{\text{trans}}$.

3. The number of tokens in place $P_{\text{os}_{-}\text{send}}$, $m_{\text{os}_{-}\text{send}}$, and the number of tokens in place $P_{\text{channel}}$, $m_{\text{channel}}$, take values of either 0 or 1. Similarly, $m_{\text{retry}}$ varies from 0 to the maximum retry limit, $\text{Retrylimit}$. This leads to a maximum of $2 \times 2 \times (\text{Retrylimit} + 1)$ states.

4. The number of tokens in place $P_{\text{os}}$, $m_{\text{os}}$ varies between 0 and $K_{\text{os}}$ and for each of these values, there could be $0 \leq m_{\text{os}_{-}\text{send}} \leq K_{\text{os}}$. This leads to $(K_{\text{os}} + 1)^2$ states due to the events at the other stations.

Using arguments (1)–(4) given above, the number of states, $N_{\text{st}}$, in the underlying CTMC of the SRN is given by

$$N_{\text{st}} \approx 4(K_{\text{os}} + 1)^2(\text{Retrylimit} + 1)(K_{\text{os}} + 1)(n + 1) \quad (B1)$$

In our model, we fix $K_{\text{os}} = 1$, $K_{\text{os}} = N - 1$ and $\text{Retrylimit} = 7$. Therefore,

$$N_{\text{st}} \approx 64N^2(n + 1) \sim O(N^2n) \quad (B2)$$

From (B2), it is observed that for $N = 10$ nodes and a back off of $n = 9$, the underlying CTMC has approximately 64,000 states. Solving the CTMC would therefore be very complex. However, this could be done using the SHARPE tool [19]. It is also observed that choosing random back-off values could result in more than $10^6$ states.

REFERENCES


**AUTHORS’ BIOGRAPHIES**

**R. Jayaparvathy** obtained her BE in Electrical and Electronics Engineering, ME in Applied Electronics from the Bharathiar University, Coimbatore, India, in 1987 and 1991, respectively. She obtained PhD in Information and Communication Engineering from Anna University, Chennai, India, in 2005. She worked as a Member Research Staff at the AU-KBC Research Center, Chennai for three years. She is currently working in Coimbatore Institute of Technology, Coimbatore as a Teaching Faculty. Her areas of research interest include Wireless Media Access Control protocols, Performance analysis, QoS and Genetic Algorithms.

**S. Anand** obtained his BE degree in Electronics and Communication Engineering, from the College of Engineering, Guindy, Anna University, Chennai, India, in 1995. He obtained his ME degree in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, India, in 1998 and his PhD degree from the Indian Institute of Science, Bangalore, India, in 2003. He is currently with Samsung India Software Operations Private Limited, Bangalore, India. His area of interests include radio resource management and quality of service in wireless networks, mobility management in heterogeneous wireless networks and medium access control in short range wireless networks. He obtained the Seshagiri Kaikkini medal for the best divisional thesis in the division of Electrical Sciences in Indian Institute of Science for the academic year 2003–2004.
S. Dharmaraja received PhD in Mathematics from Indian Institute of Technology Madras, India, in 1999 and currently working in Indian Institute of Technology, Delhi, India. Before joining this institute, he was a post-doctoral fellow at the Department of Electrical and Computer Engineering, Duke University, U.S.A. and then was a Research Associate at the TRLabs, Winnipeg, Canada. His current research interests include queuing theory, Markov modelling, performance issues of wireless networks and dependability analysis of communication systems. His work has been published in several international journals and conferences.

S. Srikanth obtained his BE degree from Anna University, Chennai, India, in 1990, MSc and PhD from University of Victoria, Canada in 1993 and 1997, respectively. He worked in Harris Corporation, Calgary, Canada from 1997 to 1999. He was a UGC Visiting Professor at College of Engineering, Anna University from 1999 to 2000. From July 2000 onwards, he is with AU-KBC Research Centre, Chennai, where he takes up R & D assignments in wireless area besides teaching. He is also a consultant to various organizations. His research interests include physical and medium access control layer techniques for wireless communications systems with emphasis on IEEE 802.11, and 802.16-based systems. He has many research papers to his credit. He has been awarded a Government of India Young Scientist Fellowship for a period of 3 years.