This is the submitted version of this conference paper:


© Copyright 2010 IEEE

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Markov Modelling of the IEEE 802.11 DCF for Real-Time Applications with Periodic Traffic

Guosong Tian and Yu-Chu Tian
Computer Science
Faculty of Science and Technology
Queensland University of Technology
Box 2434, Brisbane QLD 4001, Australia
y.tian@qut.edu.au

Abstract—Popular wireless network standards, such as IEEE 802.11/15/16, are increasingly adopted in real-time control systems. However, they are not designed for real-time applications. Therefore, the performance of such wireless networks needs to be carefully evaluated before the systems are implemented and deployed. While efforts have been made to model general wireless networks with completely random traffic generation, there is a lack of theoretical investigations into the modelling of wireless networks with periodic real-time traffic. Considering the widely used IEEE 802.11 standard, with the focus on its distributed coordination function (DCF), for soft-real-time control applications, this paper develops an analytical Markov model to quantitatively evaluate the network quality-of-service (QoS) performance in periodic real-time traffic environments. Performance indices to be evaluated include throughput capacity, transmission delay and packet loss ratio, which are crucial for real-time QoS guarantee in real-time control applications. They are derived under the critical real-time traffic condition, which is formally defined in this paper to characterize the marginal satisfaction of real-time performance constraints.

Keywords—IEEE 802.11; distributed coordination function (DCF); real-time control; modelling; performance evaluation

I. INTRODUCTION

While wireless networked control systems (WNCS) have been increasingly adopted in real-time control systems, wireless solutions and products currently available on the market, e.g., IEEE 802.11, are not designed for real-time control systems. Therefore, it is not surprising that they have some technical limitations in real-world applications. One of those technical limitations is the lack of deterministic traffic behaviors because of channel fading and external interference [1] and the random access mechanism. Another technical limitation is the inefficiency in dealing with real-time and periodic traffic in real-time control systems because the traffic behaviors in a real-time WNCS are quite different from those in normal best-effort services [2], [3].

As a popular wireless network architecture, wireless local area networks (WLANs) with the IEEE 802.11 standard have been promoted for potential industrial real-time control applications [3]. The IEEE 802.11 is adopted because of its simplicity, scalability, flexibility, and fast deployment [2], [4].

In a typical IEEE 802.11 deployment, the basic medium access control (MAC) method is the distributed coordination function (DCF), which is widely supported by most wireless product vendors. Although the point coordination function (PCF) and the IEEE 802.11e are also proposed for limited QoS support, they are not widely supported in wireless devices, e.g., access points, due to their complexity and inefficiency for normal data transmissions [5].

As the chance of collisions in wireless transmissions is low under light traffic conditions, the DCF is potentially applicable in some soft real-time control systems. Evaluation of the soft real-time performance of the DCF under non-saturation conditions has been investigated recently [2], [6]. However, to make the best use of the limited wireless resources, a real-time WNCS tends to operate each of its wireless nodes at its maximum or near maximum capacity under the constraints of the real-time requirements [7], [8]. This operating condition will be formally referred to as the critical real-time traffic condition later in Section II. Careful evaluation of the performance of the DCF under the critical real-time traffic condition would be significant for real-time QoS guarantee in real-time control systems with periodic real-time traffic. However, reports have not been found in the literature to tackle this problem. This motivates the research of this work.

Significant effort has been made in modelling of the DCF based on Markov chain theory [5], [9], [10], [11], [12], [13], [14]:

- Consideration of the DCF under saturation conditions. Bianchi et al. [9], [15] established a Markov chain model to evaluate the saturation throughput of the DCF. Following this modelling framework, Wu et al. [16] considered the finite packet retry limits in their modelling.
- Consideration of the DCF under non-saturation conditions. Ghaboosi et al. [13] modelled the network behavior of the DCF by employing the Parallel Space-Time Markov Chain. Sakurai and Wu [10] developed a stochastic model of the DCF based on a one-dimensional Markov chain for evaluating the random access delay. The packet delay performance was also evaluated by Zhai et al. [17] through employing the Markov chain modelling technique.

In addition to Markov chain theory based modelling, other modelling frameworks have also been investigated for
the DCF. Examples include stochastic reward nets based modelling [18], stochastic petri nets based modelling [19], and the modelling work by Cali et al. [20], [21].

While various traffic conditions have been considered in the literature, most of the existing work have only considered the best-effort services under saturation conditions with heavy traffic load [9], [14], [21]. In a typical example, Bianchi [9] has considered a saturation condition of the wireless traffic under the assumption that each of the wireless nodes always has packets in its transmission queue. However, this may not be the case in a real application.

To analyze non-saturation conditions in non-real-time systems, an idle state is introduced into the Markov chain modelling [22], [23], [24]. This idle state models the situation where the transmission queue becomes empty after a successful packet transmission under light traffic conditions [22]. The probability of the packet arrival into an empty queue is assumed to be constant in any time slot [22], [23]. In [24], each node is modeled as a discrete-time G/G/1 queue where the packet arrival patterns are arbitrary.

Despite the WLAN modelling progress, an analytically model has not been found in the literature to describe the behavior of the DCF with real-time constraints and periodic traffic in control applications. To satisfy the real-time performance constraints, the DCF has to work within the critical real-time traffic condition, which is typically a non-saturation condition as will be explained later in Section II. This means that empty queue states, which do not exist in Bianchi’s modelling [9], need to be embedded into the modelling framework. However, the existing models with empty queues only deal with purely random traffic generation, and thus are not applicable to the systems in which periodic traffic are generated at fixed time intervals.

This paper develops a theoretical model for evaluating the performance of the DCF based wireless networks in real-time control systems with periodic real-time traffic. The main contributions of this work include: 1) The real-time requirements and the unique feature of the periodic traffic patterns in real-time control are considered explicitly in our modelling, and the concept of the critical real-time traffic condition is formally defined to characterize marginal satisfaction of the real-time requirements; 2) The technique of the empty queue used in non-real-time systems without periodic traffic is refined for periodic real-time control tasks; and 3) Model equations are established to estimate the throughput, deadline miss ratio and average transmission delay under the critical real-time traffic condition.

The paper is organized as follows. Following this introductory section, Section II introduces some notations and definitions including the concept of the critical real-time traffic condition. Section III describes the theoretical development of an analytical Markov model for WLANs with the IEEE 802.11 DCF and periodic real-time traffic generation. Model validation is carried out in Section IV. Finally, Section V concludes the paper.

II. NOTATIONS AND DEFINITIONS

A. The Backoff Based MAC Access Scheme

The DCF implements a basic random access method, which is based on the carrier sense multiple access with collision avoidance (CSMA/CA) technique [4]. It supports two access schemes, the basic mode and the request-to-send (RTS)/clear-to-send (CTS) mode. As the basic mode is more fundamental than the RTS/CTS mode, this paper focuses on this access mode.

As in other CSMA/CA based access methods, the DCF also requires each transmitting station to sense the medium to ascertain the condition of the channel before accessing the medium. If the medium is idle, the station transmits its packet. Otherwise, it postpones its transmission until the medium is sensed free for a time interval that is the sum of a DCF Inter-Frame Space (DIFS) and the backoff interval selected randomly. It is permitted to retransmit its packet after the time interval has elapsed for the postponed packet transmission.

The length of the backoff interval is computed by the backoff time counter using the contention window mechanism. The backoff time counter is decremented as long as the channel is sensed idle. It is, however, frozen whenever a transmission is detected, and is reactivated when the channel is sensed idle again for more than a DIFS period. The station is allowed to transmit packets when the backoff time counter reaches zero.

Fig. 1 shows the basic backoff procedure of the IEEE 802.11 DCF, where several stations contend for the channel and the backoff time counters at these stations are changed according to the backoff policies. For each transmission, the backoff time is uniformly chosen in the range \([0, W-1]\), where \(W \in [W_{\text{min}}, W_{\text{max}}]\) is the backoff window size. Set \(W = W_{\text{min}}\) in any of the following cases: for the first transmission, after a successful transmission, or when the retransmission counter reaches the retry limit \(L\). After each unsuccessful transmission, \(W\) is doubled until it reaches \(W_{\text{max}}\). Once \(W\) reaches \(W_{\text{max}}\), it remains unchanged until it is reset to \(W_{\text{min}}\) after a successful transmission or the retransmission counter reaches the limit \(L\).

More specifically, \(W_j\), the value of \(W\) in the \(j\)th retry, is

\[
W_j = \begin{cases} 
2^j W_{\text{min}}, & \text{for } j = 0, 1, \ldots, M - 1, \text{ if } L < M \\
W_{\text{max}}, & \text{for } j = M, \ldots, L, \text{ if } L \geq M
\end{cases}
\]

where \(M\) is the maximum number of the stages allowed in the exponential backoff procedure. In the specifications of the IEEE 802.11b standard, the default value of \(W_{\text{max}}\) is 1024 and the recommended initial default value of \(W_{\text{min}}\) is 32 time slots.

Moreover, the DCF also supports the acknowledgment (ACK) mechanism. After receiving a frame successfully, the destination station transmits an ACK frame following a Short Inter-Frame Space (SIFS) time. If the source station does not receive the ACK within a specified ACK timeout period or detects a different frame, it schedules retransmission of the frame.

The random access properties of a WNCS network are described by two stochastic processes \(s(t)\) and \(b(t)\) for the backoff stage and the backoff counter, respectively, where \(t\)
is the time. The samples of \( s(t) \) take value from a discrete space of all possible backoff stages; while \( b(t) \) is uniformly distributed in the integer set \( \{0, 1, \ldots, W_j - 1\} \) for the \( j \)th backoff period. The pair of \( (s(t), b(t)) \) specifies the state of the backoff procedure, and its stationary distribution is denoted by

\[
b_{jk} = \lim_{t \to \infty} Pr[s(t) = j, b(t) = k].
\]

In addition, random variables used in the modelling include \( N_{idle} \) and \( N_{busy} \) for the numbers of busy and idle slots that a frame encounters during the backoff stage, respectively, and \( N_{retry} \) for the number of retries.

### B. The Critical Real-Time Traffic Condition

The period of a real-time control task is denoted as \( T \). When \( n \) wireless stations form a WNCS, they all experience packet transmission delays, which are described by a random variable \( T_{\text{delay}} \) and its mean value \( T_{\text{avg-delay}} = E(T_{\text{delay}}) \). \( T_{\text{delay}} \) is assumed to follow a Poisson distribution with the parameter \( \lambda \). The samples of \( T_{\text{delay}} \) take value from \( [T_{\text{min-delay}}, T_{\text{max-delay}}] \).

In general, reducing the control period \( T \) leads to an improvement in control performance when the WNCS is not overloaded. However, a smaller \( T \) means heavier traffic load, which normally results in higher throughput yet longer transmission delays. For real-time control applications, \( T_{\text{avg-delay}} \) should be bounded, e.g., by \( T \), and the deadline miss ratio denoted as \( R_{\text{miss-deadline}} \) should be as small as possible, i.e.,

\[
T_{\text{avg-delay}} \leq T, \quad R_{\text{miss-deadline}} \to 0.
\]  

When \( T \) is reduced to the critical value at which the real-time requirements in Eq. (3) are marginally satisfied, the network throughput reaches its maximum value (denoted by \( S \)) under the constraints. Any further reduction in \( T \) will result in dissatisfaction of Eq. (3). This critical condition is referred to as the critical real-time traffic condition in this paper.

The critical real-time traffic condition highlights the marginal satisfaction of the real-time performance requirements in Eqn (3). It is not a network saturation condition, under which the transmission queue of each transmitting station is always non-empty [15]. A station may have no packet to transmit under the critical real-time traffic condition.

To demonstrate the critical real-time traffic condition, the performance of a WLAN with 20 nodes interconnected via the 802.11 DCF is evaluated. While the detailed specifications of the WLAN will be discussed later in Section IV, Fig. 2 shows \( R_{\text{miss-deadline}}, T_{\text{avg-delay}} \) and offered traffic load versus \( T \). The real-time performance requirements are those in Eq. (3). Fig. 2 shows that when \( T \) is bigger than 33ms, the system behaves with \( R_{\text{miss-deadline}} \to 0 \) and \( T_{\text{avg-delay}} \ll T \). However, \( R_{\text{miss-deadline}} \) and \( T_{\text{avg-delay}} \) surge when \( T \) is reduced to cross the value at about 32.6ms, and Eq. (3) becomes dissatisfied though the system is still not saturated. The critical real-time traffic condition is at around \( T = 32.6 \)ms.

![Fig. 1. The backoff procedure of the IEEE 802.11 DCF.](image)

![Fig. 2. Plots of \( R_{\text{miss-deadline}}, T_{\text{avg-delay}} \) and offered traffic load versus \( T \) for a WLAN with 20 nodes (Section IV).](image)
which the channel is sensed busy because of a successful transmission. In addition, the length of a slot time is \( I_{\text{slot}} \).

Many network events occur at certain probabilities in random access networks. Let \( p \) denote the the probability that a transmitted frame collides. Other probability variables are denoted by \( p \) with subscripts: \( p_b \) for the channel being busy; \( p_{\text{incw}} \) for the channel being neither idle nor used successfully for a time slot; \( p_{\text{drop}} \) for a frame being dropped; \( p_{\text{incw}} \) for packet retries in a slot time; \( p_s \) for a successful transmission in a time slot; \( p_{\text{frame}} \) for the frame being successfully transmitted; and \( p_r \) for a station transmitting in the backoff stage during a generic slot time.

This work will introduce an empty-queue stage in the backoff procedure for real-time WNCs networks. It uses \( T_{\text{slot}} \) to denote the duration of a slot time, \( N \) to represent the number of the time slots, and \( q_k \) to describe the probability in which the system moves into the \( k \)th empty-queue state from the backoff stage, \( k \in \{0, 1, \cdots, N-1\} \).

Packet losses are characterized by the packet dropout ratio \( R_{\text{drop}} \). For delays bounded by Eq. (3), all deadline misses result from packet losses, implying \( R_{\text{drop}} = R_{\text{miss \ deadline}} \).

Table 1 tabulates the notations used in the DCF modelling of this paper.

### III. Modelling the IEEE 802.11 DCF with Periodic Traffic

Aiming to evaluate the performance of the DCF in real-time control environments, our modelling tries to capture the characteristics of the state transition of the DCF, and also to estimate the throughput and average transmission delay under the critical real-time traffic condition. Satisfaction of the critical real-time traffic condition guarantees the real-time performance requirements in Eq. (3).

The following two assumptions, common in existing Markov chain models, are also made in this work: 1) The wireless channel is in ideal conditions where the network performance is not degraded by channel conditions due to obstacles; and 2) Collisions happen with the same probability regardless of the number of retries in a backoff instance.

#### A. The Modelling Framework

Our modelling framework characterizes all possible states and their transition for the transmission process of a backoff instance in the DCF in real-time control environments. It is illustrated in Fig. 3, where the system states characterized by \( \{s(t), b(t)\} \) and their transition probabilities are clearly shown \[5, 9, 10, 11, 12, 13, 14\]. It is inspired by Bianchi’s model \[9\], but the introduction of the empty-queue stage and the consideration of the periodic traffic patterns differentiate our modelling from Bianchi’s original model.

For the non-empty queues in Fig. 3, the system state changes only when the backoff instance would be able to contend for the channel. If one or more stations transmit in a time slot, the slot is sensed busy in the backoff instance. For the subsequent time slot duering which the packet is under transmission, the backoff instance remains in the same state. The Markov process finally resumes at the first slot where the channel is again open for contention. The duration of the time slot in the backoff instances is defined in the IEEE 802.11 specifications \[4\]. The virtual time scale is discrete and integral, and a backoff slot that is open for contention triggers each “clock tick”.

As analyzed previously, in industrial control, it is not a realistic assumption that each wireless station always has a packet ready for transmission. To alleviate this unrealistic assumption, the technique of the empty-queue used in non-real-time systems without periodic traffic is refined for real-time control applications. As shown in Fig. 3, after the backoff stage, the system always transits to one of the \( N \) possible empty-queue states marked by \( j = -1 \) i.e., \( \{-1, 0\}, \{-1, 1\}, \cdots, \{-1, N-1\} \). The actual transmission delay of the current frame determines which empty-queue state the system transits to. The state \( (-1, 0) \) that waits for only one time slot \( T_{\text{slot}} \) must be experienced after every transmission. The state \( (-1, N-1) \) represents the situation that the current frame is transmitted within a single time slot \( T_{\text{slot}} \).

### TABLE I

<table>
<thead>
<tr>
<th>Notations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_{\text{incw}} )</td>
<td>Critical average real-time frequency of retries</td>
</tr>
<tr>
<td>( L )</td>
<td>Max. no. of stages in the exponential backoff procedure</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of time slots in the empty queue stage</td>
</tr>
<tr>
<td>( N_{\text{busy}} )</td>
<td>Number of transmitting stations</td>
</tr>
<tr>
<td>( N_{\text{idle}} )</td>
<td>Number of idle slots; total number of slots; and retries, respectively, which a frame encounters during backoff</td>
</tr>
<tr>
<td>( p )</td>
<td>Probability that a transmitted frame collides</td>
</tr>
<tr>
<td>( p_b )</td>
<td>Probability that the channel is busy</td>
</tr>
<tr>
<td>( p_{\text{incw}} )</td>
<td>Probability that the channel is neither idle nor used successfully for a time slot</td>
</tr>
<tr>
<td>( p_{\text{drop}} )</td>
<td>Probability that the frame is dropped</td>
</tr>
<tr>
<td>( p_{\text{incw}} )</td>
<td>Probability of the packet retries in a time slot</td>
</tr>
<tr>
<td>( p_s )</td>
<td>Probability of a successful transmission in a time slot</td>
</tr>
<tr>
<td>( p_r )</td>
<td>Probability that a station transmits in the backoff stage during a generic slot time</td>
</tr>
<tr>
<td>( q_k )</td>
<td>Transition probability in the empty-queue stage</td>
</tr>
<tr>
<td>( R_{\text{miss \ deadline}} )</td>
<td>Deadline miss ratio which is set to ( T ) in this paper</td>
</tr>
<tr>
<td>( N )</td>
<td>Critical real-time throughput</td>
</tr>
<tr>
<td>( T )</td>
<td>Transmission interval of the periodic traffic</td>
</tr>
<tr>
<td>( T_{\text{avg \ delay}} )</td>
<td>Average one-way delay</td>
</tr>
<tr>
<td>( T_{\text{ACK}} )</td>
<td>Time duration to transmit an ACK</td>
</tr>
<tr>
<td>( t_c )</td>
<td>Avg. duration during which the channel has a collision</td>
</tr>
<tr>
<td>( T_{\text{ach}} )</td>
<td>Average one-way delay of the periodic messages under the real-time saturated condition</td>
</tr>
<tr>
<td>( T_{\text{DIFS}}, T_{\text{SIFS}} )</td>
<td>DIFS and SIFS times, respectively</td>
</tr>
<tr>
<td>( T_{\text{delay}} )</td>
<td>Random variable for the one-way transmission delay</td>
</tr>
<tr>
<td>( T_{\text{LF}} )</td>
<td>Time duration to transmit the average payload</td>
</tr>
<tr>
<td>( T_{\text{LF}^{*}} )</td>
<td>Time duration to transmit a payload with length ( E(L</td>
</tr>
<tr>
<td>( T_{\text{H}} )</td>
<td>Time duration to transmit the header (including MAC and physical headers)</td>
</tr>
<tr>
<td>( T_{\text{slot}} )</td>
<td>Length of a slot time</td>
</tr>
<tr>
<td>( T_{\text{LF}} )</td>
<td>Length of average payload</td>
</tr>
<tr>
<td>( L )</td>
<td>Length of frame in a collision</td>
</tr>
<tr>
<td>( T_{\text{pdt}} )</td>
<td>Payload transmission time duration in a slot</td>
</tr>
<tr>
<td>( T_{\text{RTT}} )</td>
<td>Average round trip time</td>
</tr>
<tr>
<td>( t_s )</td>
<td>Average time that the channel is sensed busy because of a successful transmission</td>
</tr>
<tr>
<td>( T_{\text{slot}} )</td>
<td>Duration of an slot time in backoff and empty-queue (post-backoff) stage</td>
</tr>
<tr>
<td>( W_j )</td>
<td>Backoff window size in the ( j )th retry</td>
</tr>
</tbody>
</table>
Therefore, in Fig. 3, the probability $q_k = 1$ holds for only one $k$ value; for all other $k$ values $q_k = 0$. When exiting from the first empty-queue state $(-1,0)$, the station commences a new period, generates a new frame and moves to the backoff stage, enabling periodic traffic generation in a fixed time interval.

### B. Parameterization of the Modelling

The backoff stage of our model is parameterized as follows:

1) The probability $p$ that a transmitted frame collides is equal to the probability that a station senses the channel busy.

2) The states of each station are described by the pair $(j,k)$, where $j \in [0,L]$ stands for the backoff stage and $k \in [0,W_j-1]$ is the backoff delay. The state $(j,0)$, $j \in [0,L]$, represents that the packet is sent successfully after the backoff stage.

3) The value of the backoff time slot $T_{\text{slot}} = 20\mu s$ is defined in the IEEE 802.11 standard [4].

The number $N$ of the empty-queue states depends on the period $T$ and the the time slot $T_{\text{slot}}$. When setting $T = T_{\text{max\_delay}}$ that guarantees the real-time requirements in Eq. (3) for periodic messages, we have

$$N = \left\lceil \frac{T - T_{\text{min\_delay}}}{T_{\text{slot}}} \right\rceil = \left\lceil \frac{T_{\text{max\_delay}} - T_{\text{min\_delay}}}{T_{\text{slot}}} \right\rceil$$ (4)

where $\lceil \cdot \rceil$ represents the nearest rounded up integer.

The probability $q_k$ in which the system goes into the $k$th empty-queue state from the backoff stage depends on the actual transmission delay of the frame in the current period, $k \in [0,1,\ldots,N-1]$. It can be derived based on the condition $T_{\text{delay}} \in [T_{\text{min\_delay}} + jT_{\text{slot}}, T_{\text{min\_delay}} + (j+1)T_{\text{slot}})$ as follows:

$$q_k = \begin{cases} 1, & \text{for } k = j; \\ 0, & \text{otherwise}. \end{cases}$$ (5)

**Assume that $T_{\text{delay}}$ has a Poisson distribution, i.e.,**

$$Pr(T_{\text{min\_delay}} + jT_{\text{slot}} \leq T_{\text{delay}} < T_{\text{min\_delay}} + (j+1)T_{\text{slot}}) = (\lambda^j/j!e^{-\lambda}) \cdot j \in [0,N-1].$$ (6)

It follows that $T_{\text{avg\_delay}}$ can also be derived from Eq. (6) as:

$$T_{\text{avg\_delay}} = E(T_{\text{delay}}) = T_{\text{min\_delay}} + \lambda T_{\text{slot}},$$ (7)

where, the parameter $\lambda$ is to be estimated in section III-D.

For simplicity, the transmission interval of all real-time periodic traffic is set to be the same in this paper.

### C. Transition Probabilities

The non-null transition probabilities can be derived mathematically, and are summarized in Table II.

![Fig. 3. State transition diagram for real-time periodic transmissions.](image)

**TABLE II**

<table>
<thead>
<tr>
<th>Non-null Transition Probabilities of the Markov Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr((0,k)(j,0)) = (1-p)/W_{0}$, $0 \leq k \leq W_{0} - 1$ and $0 \leq j &lt; L$</td>
</tr>
<tr>
<td>$Pr((0,k)(L,0)) = 1/W_{0}$, $0 \leq k \leq W_{0} - 1$</td>
</tr>
<tr>
<td>$Pr((j,k)(j,k)) = p$, $1 \leq k \leq W_{j} - 1$ and $0 \leq j &lt; L$</td>
</tr>
<tr>
<td>$Pr((j,k)(j,k+1)) = 1 - p$, $0 \leq k \leq W_{j} - 2$ and $0 \leq j &lt; L$</td>
</tr>
<tr>
<td>$Pr((j,k)(j-1,0)) = p/W_{j}$, $1 \leq k \leq W_{j} - 1$ and $0 \leq j &lt; L$</td>
</tr>
<tr>
<td>$Pr((-1,1)(-1,1)) = 1$, $0 \leq k \leq N-1$</td>
</tr>
</tbody>
</table>

The following relations for the stationary distribution defined in Eq. (2) can be derived through chain regularities:

$$b_{-1,0} = b_{0,0},$$ (8)

$$b_{j,0} = p b_{j,0}, \quad \text{for } 0 \leq j \leq L,$$ (9)

$$b_{j,k} = \frac{W_{j-1}}{W_{j}} b_{j-1,k}, \quad \text{for } 0 \leq j \leq L, \quad 1 \leq k \leq W_{j} - 1,$$ (10)

$$b_{j,k} = \left(1 - \sum_{k=0}^{W_{j}-1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) b_{0,0}, \quad \text{for } 1 \leq k \leq N-1,$$ (11)

$$b_{j,k} = \left(1 - \sum_{k=0}^{W_{j}-1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) b_{0,0}, \quad \text{for } 1 \leq k \leq N-1,$$ (12)

$$b_{-1,k} = \left(1 - \sum_{k=0}^{N-1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) b_{1,0},$$ (13)

$$b_{j,k} = \left(1 - \sum_{k=0}^{W_{j}-1} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) b_{0,0},$$ (14)

The probability $p_r$ that a station transmits in the backoff stage during a generic slot time is given by

$$p_r = \left[1 - \left(\frac{\lambda + 1}{2}\right) b_{0,0}\right] \sum_{j=0}^{L} b_{j,0} = \left[1 - \left(\frac{\lambda + 1}{2}\right) b_{0,0}\right] \frac{1 - p^j}{1 - p} b_{0,0}.$$ (15)

A transmitted frame collides when another station also transmits during a slot time. The probability $p$ that a station senses the channel busy in a system with $n$ stations is

$$p = 1 - (1 - p_r)^n.$$ (16)

The probability $p_s$ that a successful transmission occurs in a time slot, the probability $p_b$ that the channel is busy, and
the probability \( p_c \) that the channel is neither idle nor used successfully for a time slot are respectively given by:

\[
p_s = np_s(1 - p_T)^{n-1}, \quad p_b = 1 - (1 - p_T)^{n}, \quad p_c = p_b - p_s. \tag{17}
\]

Furthermore, the probabilities \( p_{drop} \) and \( p_{suc} \) that the frame is dropped and successfully transmitted individually, respectively, are modelled as

\[
p_{drop} = p_f^j, \quad p_{suc} = 1 - p_f^j. \tag{18}\]

**D. The Critical Average Real-Time Transmission Delay \( T_d \)**

\( T_d \) is the value of \( T_{avg} = E(T_{delay}) \) under the critical real-time traffic condition. Thus, it also meets \( T_d = E(T_{delay}) \). Let us try to estimate \( E(T_{delay}) \). Appearing only in the backoff stage, \( T_{delay} \) depends on the value of a station’s backoff counter and the duration when the counter freezes.

The probability that the frame is successfully transmitted after the \( j \)th retry is \( p^j(1 - p) \). The average number of the backoff slots that a station needs to transmit a frame successfully at the \( j \)th retry is \( \sum_{j=0}^{n} (W_{b} - 1)/2 \). We have the following expressions for \( E(N_{idle}) \), \( E(N_{busy}) \) and \( E(N_{retry}) \), where \( N_{idle} \) and \( N_{busy} \) are the total numbers of the idle and busy slots that the frame encounters during the backoff stage, respectively, and \( N_{retry} \) is the number of retries:

\[
E(N_{idle}) = \sum_{j=0}^{\infty} \frac{p}{1-p} p^{j},
\]

\[
E(N_{busy}) = E(N_{idle}) \frac{p}{1-p},
\]

\[
E(N_{retry}) = \sum_{j=0}^{\infty} \frac{j p^{j}(1-p)}{1-p^j}. \tag{19}\]

For an idle slot at state \((j, k)\), a busy slot at state \((j, k)\), a failed transmission slot at state \((j, 0)\), and a successful transmission at state \((j, 0)\), the average slot lengths are \( T_{slot} = \frac{h_{slot}}{p_{h_{idle}}} + \frac{h_{slot}}{p_{h_{busy}}} T_{c} \), \((T_{c} + T_s)\), respectively. Thus,

\[
E(T_{delay}) = E(T_{idle}) T_{slot} + E(T_{busy}) \left[ \frac{p_s}{p_b} T_s + \frac{p_b - p_s}{p_b} T_c \right] + E(T_{retry}) (T_c + T_{DIFS} + T_{ACK_{timeout}}) + T_s, \tag{20}\]

where \( T_s \) and \( T_c \) are still unknown so far. If the basic access mode is used as we specified earlier, we have

\[
T_s = T_{H} + T_{E(L')} + T_{SIFS} + T_{ACK} + T_{DIFS}, \tag{21}\]

\[
T_c = T_{H} + T_{E(L')} + T_{SIFS} + T_{ACK} + T_{DIFS}. \tag{22}\]

If the size of the periodic packets is constant, i.e., \( E(LF) = E(LF^*) \), we have \( T_s = T_c \) and thus Eq. (20) is simplified to

\[
E(T_{delay}) = E(N_{idle}) T_{slot} + E(N_{busy}) T_s
+ E(T_{retry})(T_c + T_{DIFS} + T_{ACK_{timeout}}) + T_s. \tag{23}\]

In deriving Eq. (23), we have used Eq. (13).

Eq. (23) is still not solvable for \( E(T_{delay}) \) as it is related to the unknown \( \lambda \). It is noticed \( T_d \) also meets Eq. (7), which relates \( \lambda \) to \( E(T_{delay}) \) as well. It follows from Eq. (7) that

\[
E(T_{delay}) = \lambda T_{slot} + T_{min_{delay}} = \lambda T_{slot} + T_s. \tag{24}\]

Jointly solving Eqs. (23) and (24) gives \( E(T_{delay}) \) under the critical real-time traffic condition. This result is \( T_d \).

**E. The Critical Real-Time Throughput \( S \)**

The critical real-time throughput \( S \) is the maximum goodput achievable when the network with periodic transmissions meets the real-time QoS constraints in Eqn (3). It is worth mentioning that as in several other papers, e.g., [9], the concept of “throughput” used in this paper is actually the “goodput”, which measures the amount of useful information that is delivered per second to the MAC layer protocol.

The modelling of \( S \) relies on the state transition probabilities and several time duration variables. The probability that the channel is idle for a slot time is \( 1 - p_b \) and the probability that the channel is neither idle nor used successfully for a time slot is \( p_c = p_b - p_s \), as shown in Eq. (17). \( S \) satisfies

\[
S = \frac{E(T_{peak})}{E(T_{slot})} = \frac{p_s T_{E(LF)}}{(1 - p_b) T_{slot} + p_s T_c + (p_b - p_s) T_c}, \tag{25}\]

where \( T_{peak} \) and \( l_{slot} \) denote payload transmission time in a slot time and length of a slot time.

If the size of the periodic packets is constant, i.e., \( E(LF) = E(LF^*) \), we have \( T_s = T_c \) and thus Eq. (25) is simplified to

\[
S = \frac{p_s T_{E(LF)}}{(1 - p_b) T_{slot} + p_b T_c}. \tag{26}\]

It has been assumed that all \( n \) transmitting stations have the same period to send the real-time traffic periodically. The critical real-time transmission period \( T \) is computed as

\[
T = n E(LF)/S. \tag{27}\]

**F. The Critical Average Real-Time Frequency of Retries \( f_{incw} \)**

For a WLAN with \( n \) transmitting stations, the critical average real-time frequency of retries, \( f_{incw} \), is estimated from

\[
f_{incw} = n p_{incw}/T, \quad p_{incw} = \sum_{j=0}^{\infty} \frac{j p^{j}(1-p)}, \tag{28}\]

where \( T \) is computed from Eq. (27). \( f_{incw} \) is an indicator of the degree of network congestion under the critical real-time traffic condition.

**IV. Model Validation**

The developed model is validated through evaluating the performance of a WLAN, in particular under the critical real-time traffic condition. The theoretical results derived from the model are compared with those obtained from simulations using the network simulator NS2 [25]. The time span of the simulations is 50s for each of the network scenarios.

Physical specifications: Consider an IEEE 802.11b DCF based WLAN with an access point (AP) and \( n \) nodes, \( n \in \{10, 15, 20, 25, 30\} \). The AP is placed at the center of a 100m × 100m area, and the \( n \) nodes are randomly placed on a circle with the radius of 50m from the AP.

Traffic specifications: All transmitting nodes use UDP as the transport protocol, and produce periodic frames of a fixed-size of 1000 bytes payload in the same period. The traffic arrival process at each node is assumed to be Poisson distributed, i.e., Eq. (5). Both the data transmission and control rates in the
WLAN are 11Mbps. Between the basic access and RTS/CTS modes, the basic access mode is used at the MAC layer.

Parameter settings: The parameter settings used to obtain numerical results, for both the analytical model and the simulation runs, are summarized in Table III.

<table>
<thead>
<tr>
<th>Slot Time</th>
<th>20 μs</th>
<th>Packet size</th>
<th>1000bytes</th>
<th>PHY header</th>
<th>192bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFS</td>
<td>10μs</td>
<td>MAC header</td>
<td>272bits</td>
<td>ACK 112bits</td>
<td>PHY header</td>
</tr>
<tr>
<td>DIFS</td>
<td>50μs</td>
<td>Initial window size</td>
<td>32</td>
<td>Channel Bit Rate</td>
<td>11Mbps</td>
</tr>
<tr>
<td>Retry limit</td>
<td>7</td>
<td>Max. window size</td>
<td>1024</td>
<td>Control Bit Rate</td>
<td>11Mbps</td>
</tr>
</tbody>
</table>

Computation process: Firstly, derive $T_{avg\_delay}$ from Eqs. (23) and (24) under the critical real-time traffic condition; this gives the value of $T_d$. Then, $T_d$ is used to calculate all other indices including $S$, $T$, $f_{new}$ and $R_{miss\_deadline}$ under the same condition from the analytical model. In the NS2 simulation, set the period of the periodic traffic of the WLAN to be $T$ derived from the analytical model, and then simulate the WLAN and evaluate the system performance of $T_d$, $S$, $f_{new}$ and $R_{miss\_deadline}$ under the critical real-time traffic condition.

Results: Table IV shows selected numerical results. It is seen from Table IV that the analytical and NS2 simulation results of $T_d$, $S$ and $f_{new}$ match well with small relative differences (relative to the NS2 simulation results). The values of $R_{miss\_deadline}$ tend to be zero in all scenarios. These results and comparisons verify the developed model.

<table>
<thead>
<tr>
<th>The number of stations $(n)$</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical $T$ (ms)</td>
<td>15.8</td>
<td>24.1</td>
<td>32.6</td>
<td>41.0</td>
<td>49.7</td>
</tr>
<tr>
<td>Analytical $S$ (Mbps)</td>
<td>5.188</td>
<td>4.979</td>
<td>4.908</td>
<td>4.878</td>
<td>4.829</td>
</tr>
<tr>
<td>Simulation $S$ (Mbps)</td>
<td>5.187</td>
<td>4.978</td>
<td>4.906</td>
<td>4.873</td>
<td>4.824</td>
</tr>
<tr>
<td>Relative difference (%)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Analytical $T_d$ (ms)</td>
<td>2.606</td>
<td>2.898</td>
<td>3.205</td>
<td>3.496</td>
<td>3.793</td>
</tr>
<tr>
<td>NS2 simulation $T_d$ (ms)</td>
<td>2.696</td>
<td>2.937</td>
<td>3.405</td>
<td>3.545</td>
<td>3.633</td>
</tr>
<tr>
<td>Relative difference (%)</td>
<td>-3.3</td>
<td>-1.3</td>
<td>-5.9</td>
<td>-4.4</td>
<td>-4.4</td>
</tr>
<tr>
<td>Analytical $f_{new}$ (Hz)</td>
<td>81.1</td>
<td>105.7</td>
<td>127.0</td>
<td>145.2</td>
<td>161.6</td>
</tr>
<tr>
<td>NS2 simulation $f_{new}$ (Hz)</td>
<td>80.0</td>
<td>100.3</td>
<td>131.9</td>
<td>139.1</td>
<td>147.9</td>
</tr>
<tr>
<td>Relative difference (%)</td>
<td>1.4</td>
<td>5.4</td>
<td>-3.7</td>
<td>4.4</td>
<td>9.3</td>
</tr>
<tr>
<td>Analytical $R_{miss_deadline}$ (%)</td>
<td>2.4E-6</td>
<td>1.2E-4</td>
<td>3.7E-4</td>
<td>8.1E-4</td>
<td>1.5E-3</td>
</tr>
<tr>
<td>NS2 simulation $R_{miss_deadline}$ (%)</td>
<td>0.24</td>
<td>0.18</td>
<td>0.29</td>
<td>1.3</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table IV also shows that as $n$ increases, $T_d$ and $f_{new}$ increase whereas $S$ decreases. This indicates that with the increase in the number of nodes the overall network performance deteriorates due to the contentions among the nodes with the same access probability. The increases in $T_d$ and $f_{new}$ lead to degradation of the real-time performance of the system.

To give a clear view of the critical real-time traffic condition, Fig. 2 in Section II shows the simulation results for the WLAN with 20 stations ($n = 20$). It is seen from Fig. 2 that reducing $T$ that leads to increased traffic load has limited effect on the real-time performance as long as $T$ is bigger than about 33ms, as evidenced by $T_{avg\_delay} < T$ and $R_{miss\_deadline} \rightarrow 0$. However, when $T$ is further reduced to cross the critical value at about $T = 32.6$ms, $T_{avg\_delay}$ and $R_{miss\_deadline}$ increase dramatically, resulting in dissatisfaction of the real-time requirements in Eq. (3). Corresponding to the critical real-time traffic condition, this critical value of $T = 32.6$ms obtained from the NS2 simulation matches well with that derived from the analytical model (Table IV).

V. Conclusion

An analytical Markov model has been developed to describe the network dynamics of the IEEE 802.11 DCF for real-time applications with periodic real-time traffic generation. Corresponding to the best achievable real-time QoS performance, the critical real-time traffic condition has also been formally defined to capture the marginal satisfaction of real-time performance constraints. Different from existing models that deal with purely random traffic generation, the model developed in this paper characterizes the periodic traffic generation, a unique feature of real-time control systems. With this modelling, network QoS performance is evaluated quantitatively under the critical real-time traffic condition. The Markov modelling has been validated through numerical examples.

Acknowledgments

This work was supported in part by the Australian Government’s Department of Innovation, Industry, Science and Research (DIISR) for its support under International Science Linkages (ISL) Grant Number CH070083, and by the Natural Science Foundation of China (NSFC) under International Collaboration Grant Number 40740420661.

References


