Accurate QoS Prediction for CSMA/CA Systems with Uncorrelated Interference

Henning F. Schepker and Ahmad Saad
Fraunhofer Institute for Embedded Systems and Communication Technologies ESK, Munich, Germany
{henning.schepker, ahmad.saad}@esk.fraunhofer.de

***Abstract*—Coexistence of wireless systems in unlicensed bands is considered a severe performance bottleneck, given the heterogeneous and uncoordinated nature of the wireless technologies. A promising approach to address this issue is to apply cognitive radio (CR) techniques, which are capable of accurately predicting the quality-of-service (QoS). This enables highly reliable QoS management and performance guarantees for applications with strict requirements, such as industrial automation or connected driving. Furthermore, accurate QoS prediction is very important for the reliability of safety critical applications. To this end, we present a novel analytical model for predicting the probability distribution of the latency based on Markov Chains (MC) for transmission systems, which employ Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) as a medium access scheme. Further, we validate the high accuracy of the prediction model using ns-3 simulations of an IEEE 802.11n communication.**

***Index Terms*—CSMA/CA, cognitive radio, QoS prediction, Markov chains, IEEE 802.11n**

I. Introduction

In recent years, the number of applications for wireless communication with strict Quality-of-Service (QoS) requirements, such as wireless factory and connected driving, has steadily increased. However, one major problem is the lack of guarantees regarding QoS. This problem is even more pronounced in unlicensed bands, such as industrial, scientific and medical (ISM) bands, where an increasing number of heterogeneous communication systems need to share the same spectrum. In this context, the lack of coexistence mechanisms leads to a severe performance bottleneck and makes the transmission unreliable. Thus, there is a need for coexistence in a shared spectrum while fulfilling QoS requirements, especially for applications with strict requirements regarding latency or reliability. For example, in the context of industrial communication, different technologies such as those based on IEEE 802.11 and those based on IEEE 802.15.4 need to coexist in the 2.4 GHz band [1]–[3], while in the context of connected driving the coexistence between IEEE 802.11p and LTE-V2X in the 5.9 GHz band is currently being investigated [4].

A promising approach to address the coexistence problem is using cognitive radio (CR) techniques. Employing CR enables a transmitter to adapt its transmission dynamically in order to maintain a high QoS, while at the same time peacefully coexisting with other wireless technologies, e.g., see [5]. To this end, a CR system has to view every other transmission technology as a primary user (PU), since they do not interact with the CR or adhere to its technology [6]. In order to enable reliable CR systems, we define an accurate QoS prediction model, based on accurate computation of the probability distribution of the latency under the influence of an uncorrelated PU. More specifically, in order to define an accurate QoS prediction we employ a Markov Chain (MC) to model the probability distribution of the latency resulting from both the contention processes as well as the retransmissions of an IEEE 802.11n system [7]. In addition to CR techniques, this QoS prediction can also be used to make safety critical applications more reliable.

Previous works, such as [8]–[13], have considered various aspects of the expected latency or delay of Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) systems in various scenarios. For example, the seminal work of Bianchi [8] focuses on deriving a model for the throughput of a CSMA/CA system for a varying number of stations. Other related works focus on different aspects of the contention process, such as influence of multi-hop communication [10], [12], vehicular communication [9], and influence of MIMO as well as varying number of stations [11]. However, these papers have in common that they only consider cases where all transmissions in the environment are from CSMA/CA stations. Thus, the main contribution of this paper is that for the first time we consider the influence of a heterogeneous uncorrelated transmission system interfering with the CSMA/CA system in a coexistence scenario. While MCs and probability distributions have been studied for the contention process in previous works, e.g., [8], [13], here we define the precise probability distribution for the total delay including contention and retransmission processes.

This paper is organized as follows: First, we introduce the proposed CR system for QoS management. Afterwards, we define and explain the analytical model for the probability distribution of the delays. For this model, we first introduce a Markov Chain for the delay, subsequently derive the corresponding delay distribution and then define QoS prediction in terms of given QoS metrics based on this model. Finally, we verify the prediction model by comparing it with ns-3 simulations of an IEEE 802.11n system, and investigate the influence of the PU on the QoS metrics.

II. System Description

In this section, we provide a description of the transmission scenario, as well as propose a cognitive radio (CR) system, which exploits the QoS prediction for coexistence management. We consider a scenario where one station of our CR system communicates with the corresponding access point, while an independent heterogeneous transmission system is operating in same spectral part of an ISM band, i.e., using the same channel. While having a single CR station might sound like a simplification, we can view this station as the aggregated demand of multiple CR stations. In an ISM band, all devices can be considered contending primary users (PUs) which need to be protected [6], and thus an interfering heterogeneous transmission system has to be viewed as a PU.

We assume the PU is independent of any transmissions of the CR system. More precisely, even though the transmission of the PU interferes with the CR communication, we assume that the PU is neither aware of the CR transmissions, nor does it adapt its behavior based on transmissions of the CR system, which is a widely adopted assumption, e.g., [14]. Further, we assume that the PU is a stationary process, such that its behavior does not change over time.

For the CR system, we assume Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) medium access is used, as implemented in IEEE 802.11n [7]. While we focus on 802.11n as a common implementation of a CSMA/CA medium access, the results of this paper can easily be adapted to different implementations of CSMA/CA. Further, we assume a saturated model for the CR system [8], i.e., whenever a packet was successfully transmitted the next packet is already waiting in the queue.

In order to consider a challenging worst-case scenario, we assume uncorrelated PU medium access in the time domain, which is slot synchronous with the CR system for simplicity. More specifically, we assume that during any protocol slot of the CSMA/CA system the PU is independent identical distributed (i.i.d.) active with a probability of p_{on}, i.e., its duty cycle. Consequently, the PU is not active with a probability of $p_{\text{off}} = 1 - p_{\text{on}}$. Due to the assumption of a stationary process for the PU, p_{on} and p_{off} are both constant over time. While any possible activity pattern can occur using this model, such as bursts of various length, it does not assume any structure regarding PU activity. Further, we assume that the interference from the PU is sufficiently strong that whenever there is a collision of a transmitted packet with the PU, the entire packet is lost and has to be retransmitted. This means that we use a worst-case assumption for the transmission, since in practice the channel code should be able to correct some errors.

We show the architecture of an appropriate CR transmitter in Figure 1. The CR is assumed to have two radio frequency (RF) interfaces, one connected to the "Sensing" block, which is used for spectrum sensing, and the other connected to "Tx/Rx" block, which is used for data transmission and reception. The "Sensing" block scans N channels, each with an index $n \in \{1, \ldots, N\}$, and forwards the sensing output per channel to the corresponding QoS prediction block labeled with "$p(\tau > \epsilon|n)$". Each of these QoS prediction blocks uses the measured information to estimate the QoS metric for one channel. In Figure 1, we use probability $p(\tau > \epsilon|n)$ that the

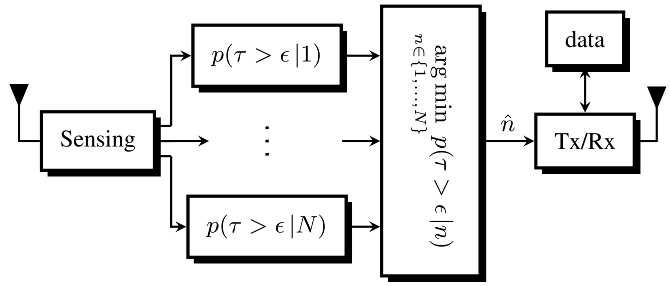


Figure 1. Architecture of a CR system employing QoS-prediction to select the best channel.

delay τ is above a threshold ϵ given the choice of channel n as a QoS metric. This can be replaced by another QoS metric according to the requirements of the application, as will be discussed in Section III-C. Subsequently, an optimization block selects the channel $\hat{n}$, which results in the best QoS, i.e., here smallest delay violation probability, and forwards this channel index to the "Tx/Rx" block. The "Tx/Rx" block then in turn uses the selected channel $\hat{n}$ for transmission of the data of the application layer.

Since this CR system selects channels based on a best effort approach, it will not be able to achieve the required QoS in cases where no such channel exists. However, due to the highly accurate QoS prediction the CR system is able to send an appropriate warning to the higher layers, which can employ appropriate countermeasures, such as switching to a different transmission technology. This further improves reliability and is particularly suitable in safety critical applications.

III. QoS Analytical Model

In this section, we define an analytical model, which enables precise computation of the delay caused by a successful CSMA/CA transmission in the presence of a PU. In order to explain this model better, we first define a Markov Chain (MC) describing the contention and retransmission processes resulting in this delay. Afterwards, we derive the analytical model corresponding to the MC. Finally, we discuss QoS metrics based on this model.

A. Markov Chain

For the MC, we model the process from a new packet being ready to send until either this package is transmitted successfully or it is dropped. In this MC we associate each possible transition between states with both a probability p and a delay τ. In cases of a forced transition, i.e., $p = 1$, the probability is omitted, and similarly the delay is omitted in cases where a transition does not increase the delay, i.e., $\tau = 0$. Therefore, the resulting delay is determined by sum of the delays associated with all transitions until reaching a final state.

The MC in Figure 2 models three separate processes, which all contribute to the overall delay: The Inter-Frame Spacing (IFS), contention and retransmission processes. The overall structure of the MC is defined by the retransmission process, which employs the contention process prior to each

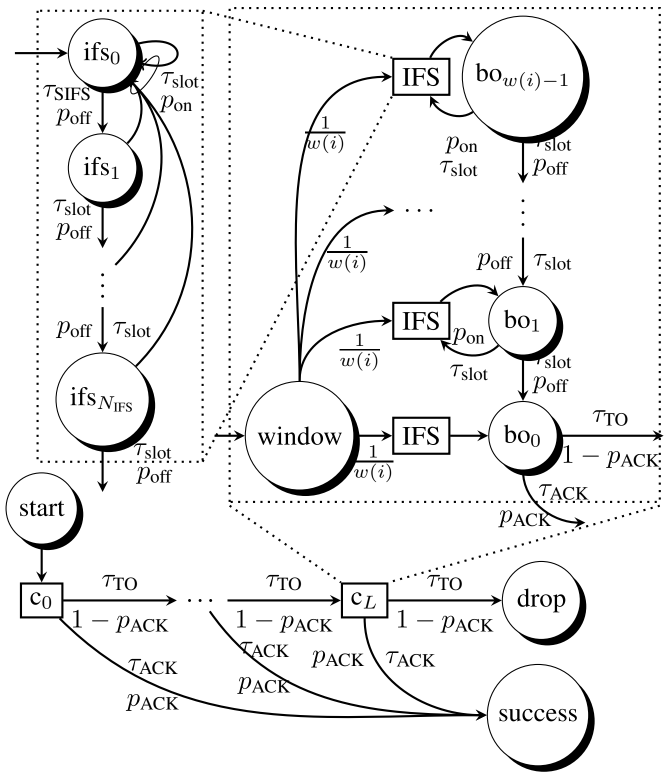


Figure 2. Markov chain for the delay of a CSMA/CA transmission under the presence of a PU.

retransmission, and the contention in turn may employ the IFS process multiple times. For visual clarity, we have summarized the IFS and contention processes into sub-chains, and only exemplary show the structure of one instance of both sub-chains in Figure 2. We depict a state with a circle and a sub-chain, which contains multiple states and transitions, with a rectangle.

The retransmission process retransmits the packet until either an acknowledgment is received or the packet is dropped. Thus, this process is mainly determined by the probability of a successful transmission p_{ACK}. Since the PU is a stationary process, p_{ACK} is constant for all retransmission. We assume that the delay caused by transmitting a single packet of a given size over the air and receiving the acknowledgment is constant and denote it with τ_{ACK}, as given by

$$\tau_{\text{ACK}} = 2 \cdot \tau_{\text{prop}} + \tau_{\text{SIFS}} + \tau_{\text{tx,packet}} + \tau_{\text{tx,ACK}} \,, \qquad (1)$$

where τ_{prop} is the propagation delay, τ_{SIFS} is the length of the Short Inter-Frame Space (SIFS), and $\tau_{\text{tx,packet}}$ and $\tau_{\text{tx,ACK}}$ are the time durations for transmitting the packet and the acknowledgment respectively. The timeout τ_{TO} for waiting for an acknowledgment, i.e., the delay of a failed transmission, will in general be slightly larger than τ_{ACK}, i.e., $\tau_{\text{TO}} \geq \tau_{\text{ACK}}$, in order to account for unknown propagation delays [7]. If the i^{th} retransmission is successful, the MC transitions into the "success" state and terminates. Here, in order to make the nomenclature more concise, we treat the initial transmission as the 0^{th} retransmission. In case the i^{th} retransmission fails, i.e., with probability $1 - p_{\text{ACK}}$, the station retransmits the packet until the L^{th} retransmission. If this retransmission was still not successful, the packet is dropped, as modeled by a transition to the "drop" state.

Prior to the i^{th} retransmission the contention process is performed, as represented by the sub-chains labeled "c$_i$". This process is determined by initializing the back-off counter to a number in the range of 0 to $w(i)-1$ uniformly at random, and decrementing this counter for each free slot until it reaches 0 [7]. Here, $w(i)$ is the contention window size of the i^{th} retransmission: For the 0^{th} retransmission $w(i)$ is set to w_{min} and is doubled for each subsequent retransmission, up to a maximum of w_{max}. Thus, $w(i)$ is given by

$$w(i) = \begin{cases} 2^i w_{\text{min}} & 2^i w_{\text{min}} < w_{\text{max}} \\ w_{\text{max}} & \text{otherwise} \end{cases} \,. \qquad (2)$$

In the MC, randomly initializing the back-off counter corresponds to a transition from the "window" state to a random back-off state "bo$_n$", with $n \in \{0, \ldots, w(i)-1\}$. Each of those transitions has the same probability $1/w(i)$ to model uniform random behavior. The index of each back-off state is used to indicate which value of the back-off counter this state corresponds to. It should be noted that there is no direct transition from "window" to "bo$_n$", instead first the IFS sub-chain has to be traversed, as explained later on. If the medium is perceived as free for the duration of a slot τ_{slot} [7], i.e., with a probability of p_{off}, the contention process steps to the next back-off state. Once the contention process has reached state "bo$_0$", the contention is finished and the node performs the i^{th} retransmission, which is modeled by exiting the "c$_i$" sub-chain.

Whenever the PU is active during the contention process, the transmitter first has to wait for free medium with a duration of the appropriate IFS [7] before continuing, as depicted by the "IFS" sub-chain. In most cases, the IFS will be either the Distributed Inter-Frame Spacing (DIFS) or the Arbitration Inter-Frame Spacing (AIFS). The type of IFS used determines the number of nodes in the "IFS" sub-chain, similar to $w(i)$ for the contention process. This is because all types of IFS share the property that the duration of the IFS is given by $\tau_{\text{SIFS}} + N_{\text{IFS}}\tau_{\text{slot}}$ with $N_{\text{IFS}} \in \mathbb{N}$ [7]. The structure of the IFS process is similar to the contention process, with the major difference being that whenever the PU is active during the IFS process, the process has to restart at the beginning and wait for free medium of the entire IFS duration. Due to the assumption of a saturated model, the channel is busy prior to contention, and thus this IFS process also has to be performed at the start of each contention process.

B. Delay Distribution

In this section, we derive the probability distribution of the total delay until the transmission is successful based on the MC defined in Section III-A. To simplify the model, we quantize the distribution to be discrete in time, e.g., number of microseconds, and thus compute the distribution as a Probability Mass Function (PMF). In this paper, we denote a PMF with $f(k)$ and assume that all PMFs have a positive support, i.e., $\sum_{k=-\infty}^{\infty} f(k) = 1$ and $f(k) = 0 \, \forall k < 0$ holds for all $f(k)$, with $k \in \mathbb{Z}$.

In the following, we will use the stochastic property that if a random variable Z is the sum of two independent random variables X and Y, with PMFs $f_X(k)$ and $f_Y(k)$ respectively, then the PMF $f_Z(k)$ of Z is given by

$$f_Z(k) = f_X(k) * f_Y(k)\,, \tag{3}$$

where "$*$" is the discrete convolution operator. Further, we will use the property that if we add a constant value τ to a random variable X with PMF $f(k)$, then the resulting random variable will have a PMF given by $f(k) * \delta(k-\tau)$, where $\delta(k)$ is the discrete dirac-function given by

$$\delta(k) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}\,. \tag{4}$$

Finally, if a random variable X has a PMF of $f(k)$, then the PMF for the sum over n i.i.d. random variables X is given by

$$f(k)^{*n} = \underbrace{f(k) * f(k) * \cdots * f(k)}_{n} \text{ with } f(k)^{*0} = \delta(k)\,. \tag{5}$$

We define the overall delay in three steps. First, we define the delay caused by the IFS process. Secondly, we define the delay of the contention process based on the delay distribution of the IFS process. Finally, we use the delay distribution of the contention process to define the overall delay of a successful transmission.

The IFS process has to wait for consecutive free medium with duration of a least $\tau_{\text{SIFS}} + N_{\text{IFS}}\tau_{\text{slot}}$. Thus, whenever the medium is busy within this duration, the process has to restart. Due to this restarting of the process, we write the resulting delay as a series of PMFs $f_{\text{ifs},m}(k)$, with $m \in \{0, \ldots, N_{\text{IFS}}\}$. For each step of the series, the delay is primarily defined by the delay caused by all previous steps $f_{\text{ifs},m-1}(k)$. The delay of the previous steps has to occur at least once, but can occur multiple times due to restarting the process. Since both an active PU or one free a slot increase the delay by one slot, this duration is added to the previous delays regardless of whether the IFS process restarts or not. Each restart occurs with probability of p_{on}, and thus multiple consecutive restarts without progressing to the next step become increasingly unlikely. The entire PMF is then scaled by p_{off} as this the probability of proceeding to the next step. Therefore, the series of PMFs $f_{\text{ifs},m}(k)$ is given by

$$f_{\text{ifs},m}(k) = p_{\text{off}} \cdot \sum_{n=1}^{\infty} p_{\text{on}}^{n-1} \left[\delta(k-\tau_{\text{slot}}) * f_{\text{ifs},m-1}(k)\right]^{*n}\,, \tag{6}$$

where $f_{\text{ifs},0}(k)$ is given by

$$f_{\text{ifs},0}(k) = p_{\text{off}} \cdot \delta(k-\tau_{\text{SIFS}}) * \sum_{n=0}^{\infty} \left[p_{\text{on}} \cdot \delta(k-\tau_{\text{slot}})\right]^{*n}\,. \tag{7}$$

In this context, $f_{\text{ifs},0}$ is a special case, since there is no previous state and also we set the first step to be τ_{SIFS}. Here, we assume that this does not affect our assumption of slot synchronicity, even though $\tau_{\text{SIFS}} \neq \tau_{\text{slot}}$, and will show in Section IV-B that this only has a very small influence on the results.

In order to define the distribution for the contention process, we first define the distribution of the delay caused by transitioning through a single back-off state "bo$_n$". Here, the delay distribution consists of two parts, one for exiting the state and one for waiting in the state, i.e., the transition to the "IFS" sub-chain and exiting this sub-chain. In order to exit a back-off state, the medium has to be perceived as free for the duration of a slot. This occurs with a probability of p_{off} and incurs a delay of τ_{slot}. Prior to leaving the state there can be an arbitrary number of back-off steps, i.e., freezing the back-off counter, due to perceiving an occupied medium. Such a back-off step occurs with a probability of p_{on} and causes a delay of one slot plus the delay of the IFS process. Using Equation (6), the PMF of the delay caused by a back-off state is given by

$$\begin{aligned} f_{\text{bo}}(k) = p_{\text{off}} \cdot \delta(k-\tau_{\text{slot}}) * \sum_{n=0}^{\infty} \Big[& p_{\text{on}} \cdot f_{\text{ifs},N_{\text{IFS}}}(k) \\ & * \delta(k-\tau_{\text{slot}}) \Big]^{*n}\,. \end{aligned} \tag{8}$$

Next, we define the delay caused by the contention process prior to the i^{th} retransmission. To this end, we compute the delay caused by transitioning through n back-off states. Using Equation (5) and Equation (8) this delay is given by $f_{\text{bo}}(k)^{*n}$. As discussed in Section III-A, each value of n has the same probability of $1/w(i)$. Thus, the PMF of the contention window is given by the sum of visiting n back-off states for all values of $n \in \{0, \ldots, w(i)-1\}$ weighted with their individual probability of $1/w(i)$. Further, we add the delay of the IFS, as this process is performed once at the start of each contention, see Section III-A. Thus, the PMF of the delay caused by the contention prior to the i^{th} retransmission is given by

$$f_{\text{c},i}(k) = \frac{1}{w(i)} f_{\text{ifs},N_{\text{IFS}}}(k) * \sum_{n=0}^{w(i)-1} f_{\text{bo}}(k)^{*n}\,. \tag{9}$$

In order to define the total delay of a successful packet, we first need to define the distributions of delays caused by successful and failed retransmissions. As discussed in Section II, we use the pessimistic assumption that whenever the PU is active during the duration of τ_{ACK}, the transmission has failed and the packet has to be retransmitted. Thus, we require free medium for the entire duration of the transmission, in order to transmit successfully. Thus, a successful transmission adds a delay of τ_{ACK} and this event occurs with a probability of p_{ACK}. Based on Equation (1), this acknowledgment probability is given by

$$p_{\text{ACK}} = p_{\text{off}}^{N_{\text{TX}}}\,, \tag{10}$$

where N_{TX} is the whole number of protocol slots required for transmission of the packet and receiving the acknowledgment, in other words τ_{ACK} converted into multiples of slots τ_{slot}. In the case of an incorrect transmission, i.e., no

ACK is received, the transmitter waits until a timeout duration τ_{TO} has passed before retransmitting the packet. Thus, using Equation (3) the delay PMF of i failed prior transmissions is given $(1-p_{\text{ACK}})^i \delta(k-\tau_{\text{TO}})^{*i}$, where the probability of failing i consecutive retransmissions is given by $(1-p_{\text{ACK}})^i$. Consequently, the case where the packet could not be successfully delivered within L retransmissions and it is dropped occurs with a probability of

$$p_{\text{drop}} = (1-p_{\text{ACK}})^{L+1} \,. \tag{11}$$

Prior to each transmission attempt, the contention process has to be performed. Using Equation (3) the PMF of the delays caused by all contention processes including the contention prior to the i^{th} retransmission is given by

$$f_{\mathcal{C}}(k,i) = f_{\text{c},0}(k) * f_{\text{c},1}(k) * \cdots * f_{\text{c},i}(k) \,. \tag{12}$$

Therefore, the total delay caused by a successful delivery of the packet within L retransmission is given by the sum of all delays caused by previous failed retransmissions, the successful retransmission and their corresponding contention processes. Thus, the overall delay PMF is given by

$$\begin{aligned} f_{\text{delay}}(k) = \frac{p_{\text{ACK}} \cdot \delta(k-\tau_{\text{ACK}})}{1-p_{\text{drop}}} * \sum_{i=0}^{L} (1-p_{\text{ACK}})^i \\ \cdot\, \delta(k-\tau_{\text{TO}})^{*i} * f_{\mathcal{C}}(k,i) \end{aligned} \,. \tag{13}$$

C. QoS Metrics

In the following, we discuss the implications of the delay PMF for successful transmission $f_{\text{delay}}(k)$ for various QoS requirements. Two common QoS requirements are high reliability, e.g., low packet loss rate, and a low latency, i.e., a sufficiently high probability of the total delay being below a given threshold. For both of these requirements appropriate metrics can easily be derived from the analytical model: On the one hand, the packet loss rate is given by the drop probability in Equation (11). On the other hand, the probability of violating the delay requirement of a given delay τ can easily be computed using Equation (13) as

$$p(\tau > \epsilon) = \sum_{k=\epsilon+1}^{\infty} f_{\text{delay}}(k) \,. \tag{14}$$

Here, we omit the channel index n used in Figure 1, as both the analytical model and the QoS metrics focus on the current channel.

The analytical model derived in Section III-B can also be used to derive other QoS metrics as well. For example, the throughput is given by

$$T = N_{\text{payload}} / E\{f_{\text{delay}}(k)\} \,, \tag{15}$$

where N_{payload} is the size of the payload and $E\{f_{\text{delay}}(k)\}$ is the expected value of the delay. Additionally, the packet error rate is directly given by the analytical model as $1-p_{\text{ACK}}$.

While computing the precise PMF would be too time-consuming for most in real-time systems, a given QoS metric for a target system primarily depends on the parameter p_{on}. This enables the QoS metric to be pre-computed and stored in a lookup table, making computational cost feasible for real-time systems. If other parameters also change during runtime, such as payload size, a separate lookup table could be computed for each parameter set.

Description	Value
Minimum window size	$w_{\text{min}} = 16$
Maximum window size	$w_{\text{max}} = 1024$
Maximum retransmissions	$L = 7$
Slot length	$\tau_{\text{slot}} = 9\ \mu s$
SIFS	$\tau_{\text{SIFS}} = 10\ \mu s$
IFS slots	$N_{\text{IFS}} = 3$
Modulation and Coding Scheme	MCS = 3
Packet & ACK transmission delay	$\tau_{\text{ACK}} = 400\ \mu s$
ACK timeout	$\tau_{\text{TO}} = 401\ \mu s$
Packet & ACK transmission slots	$N_{\text{TX}} = 41.\overline{4}$

Table I
SUMMARY OF SIMULATION PARAMETERS.

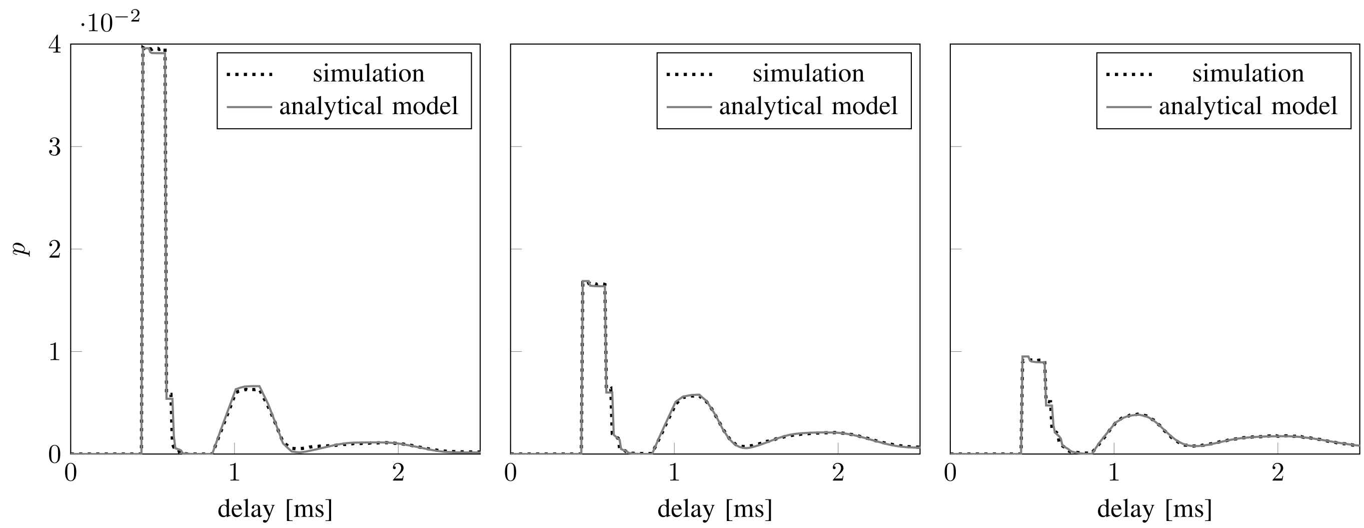
IV. VERIFICATION

In this section, we verify the analytical model by comparing it with simulation results. As an exemplary technology, we consider the IEEE 802.11n standard [7] and for simplicity assume that Multiple-Input and Multiple-Output (MIMO) transmission is not used, i.e., only a single antenna is used for transmission, as depicted in Figure 1. The appropriate parameters of the model introduced in Section III for this technology choice are given in Table I [7]. For the PHY-Layer, we assume Modulation and Coding Scheme (MCS) 3 [7], i.e., 16-QAM and code rate $1/2$. Further, we set the propagation delay to be $1\ \mu s$. Therefore, for a packet size of 1000 Bytes, the total transmission delay including acknowledgment is $\tau_{\text{ACK}} = 400\ \mu s$ and this transmission occupies $N_{\text{TX}} = 41.\overline{4}$ protocol slots. For the timeout we use $\tau_{\text{TO}} = 401\ \mu s$. As an IFS, we use the AIFS with "Access Category Best Effort", and thus the AIFS has a duration of $N_{\text{IFS}} = 3$ slots in addition to the SIFS.

For the analytical model, we compute the model as being discrete in microseconds, i.e., the parameter k of $f_{\text{delay}}(k)$ represents a number of microseconds. Since not all timings are a whole multiple of the slot duration τ_{slot}, e.g., $\tau_{\text{SIFS}}/\tau_{\text{slot}} = 1.\overline{1}$, microseconds are better suited in this case. For visual clarity of the results of both our analytical model and the simulation, we quantize all delays to the nearest slot, and show the delay distribution as being discrete in number of slots.

A. Simulation setup

For the simulation results, we use the "ns-3" [15] simulation environment (version 3.28), since it is commonly used and has many components available for simulating IEEE 802.11n. More specifically, we use the "SpectrumWifiPhy" and "AdhocWifiMac" models of ns-3 [15]. We defined two WLAN stations adhering to the 802.11n standard, one acting as a source and the other as a sink. Whenever there is no active



a) $p_{\text{on}} = 0.01$ b) $p_{\text{on}} = 0.03$ c) $p_{\text{on}} = 0.05$

Figure 3. Delay distribution for $p_{\text{on}} = 0.01$, $p_{\text{on}} = 0.03$ and $p_{\text{on}} = 0.05$.

packet, the source transmits a new packet of a fixed size and the sink transmits the ACK once a packet is received. We set the channel between the source and the sink to be ideal, i.e., the signal is received exactly as it is transmitted, as this is a common approach in MAC layer design analysis [8], and allows us to focus on the influence of the PU.

For the PU, we use a waveform generator, which generates a broadband signal occupying the channel of the WLAN transmission. To ensure that a packet is dropped whenever a collision occurs and to make sure that the PU is recognized as an occupied medium, we set the interference power to be much higher than the detection threshold. The PU waveform is generated asynchronous to the IEEE 802.11n transmission, but uses the slot length for active and inactive periods.

B. Numerical Results

In Figure 3, we compare the delay distributions computed with the analytical model to the empirical delay distributions measured in the ns-3 simulation for three different activity probabilities $p_{\text{on}} = 0.01$, $p_{\text{on}} = 0.03$ and $p_{\text{on}} = 0.05$. First, we consider probabilities of delays for cases where the initial transmission is successful, as shown by the first plateau of non-zero probabilities. Here we notice that the values of these probabilities decrease for increasing p_{on}. This is because an increased p_{on} also decreases p_{ACK}, i.e., the probability of succeeding in the first attempt. Secondly, for increasing p_{on} the partial distributions corresponding to the individual retransmissions become broader. This shows the effect of p_{on} on the contention process: A larger p_{on} increases the probability that a back-off and IFS process has to be performed, which makes larger delays more likely.

When comparing the simulation with the analytical model in Figure 3, it is clear that the simulation results and the analytical model match with high accuracy, with two minor differences: On the one hand, the shape of the first plateau differs slightly, which is due to differences in the IFS process between analytical model and simulation. On the other hand, there is a slight shift in the distribution for the delays of the retransmissions, especially for smaller p_{on}. This is due to the fact that analytical model assumes slot synchronicity, while the interferer is asynchronous in the ns-3 simulation. These results show that there is only a small error due to the assumption of slot synchronicity.

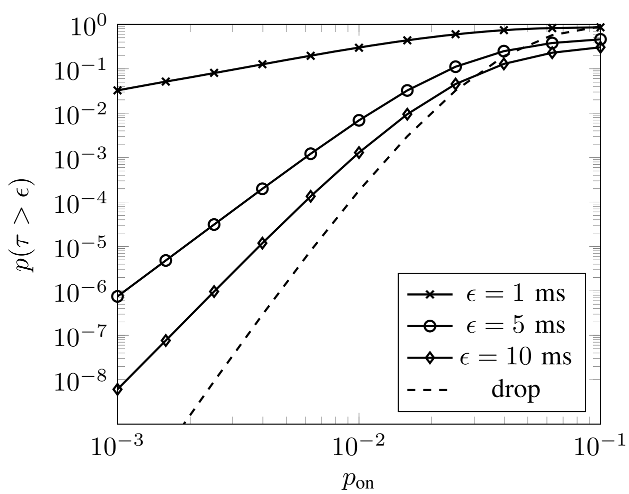


Figure 4. Probability of delay for successful transmission exceeding τ and drop probability for a range of p_{on}, for packet size of 1000 Bytes.

Additionally, we have performed the χ^2-test to measure goodness-of-fit. The resulting p-values were $p(p_{\text{on}} = 0.01) \approx 0.9910$, $p(p_{\text{on}} = 0.03) \approx 0.9879$ and $p(p_{\text{on}} = 0.05) \approx 0.9961$. This indicates a very accurate fit between the analytical model and the ns-3 simulation.

Figure 4 shows the probabilities of violating the delay threshold $p(\tau > \epsilon)$, as defined in Section III-C, for several delay requirements ϵ, as well as p_{drop} across several p_{on} and for a packet size of 1000 Bytes. From the results in Figure 4 it is clear that for very strict delay requirements,

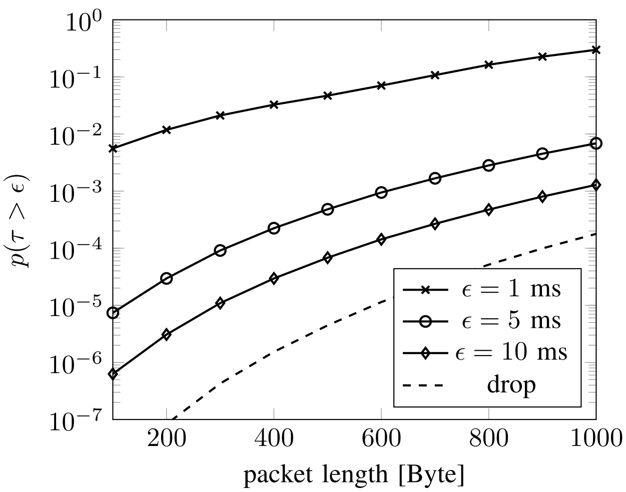


Figure 5. Probability of delay for successful transmission exceeding a QoS requirement and drop probability for different packet lengths and $p_{\text{on}} = 0.01$.

i.e., $\epsilon = 1$ ms, there is a high chance of failure, and decreasing p_{on} only slightly decreases the probability of violating this requirement. However, for less strict requirements, i.e., $\epsilon = 5$ ms and $\epsilon = 10$ ms, we observe that the violation probability significantly improves for decreasing p_{on}. For a high activity probability of $p_{\text{on}} = 0.1$, Figure 4 shows that almost all packets are dropped, which makes the transmission unusable for practical purposes. This is due to the assumption, that the transmission fails whenever the PU is active during this transmission, as discussed in Section III-A. However, p_{drop} improves significantly for smaller p_{on}. So much so that for $p_{\text{on}} = 0.001$ the drop probability is $p_{\text{drop}} \approx 7.4 \cdot 10^{-12}$, which allows for highly reliable communication.

Finally, Figure 5 shows the influence of the packet length on the delay violation probability $p(\tau > \epsilon)$ and the drop probability p_{drop} for $p_{\text{on}} = 0.01$. Here, we see that decreasing the packet size significantly improves all QoS metrics. This is because a smaller packet size results in a shorter τ_{ACK}, see Equation (1), and thus an increased p_{ACK}, see Equation (10). This in turn increases the expected number of necessary retransmissions. Since the packet size only affects the retransmissions, we see that the packet size has an almost identical effect on all QoS metrics, as opposed to the results in Figure 4.

V. Conclusion

In this paper, we presented an accurate analytical model for the probability distribution of the delay of a CSMA/CA based transmission in the presence of an uncorrelated PU or interferer, in order to enable precise QoS prediction. To this end, we have defined a Markov Chain describing the behavior of the transmitter and the influence of the PU. We have compared this analytical model with a packet-based IEEE 802.11n simulation, and have shown a very good match of the results. Further, we have investigated the influence of the PU on the prediction of various QoS metrics. This QoS prediction enables coexistence or QoS management using cognitive radio mechanisms, as well as enabling a signaling of the expected QoS to higher layers, which allows the application to adapt accordingly and enables safety critical applications.

While our PU model contains all possible activity patterns, most real life PUs will have some structure for their activity and inactivity, such as periodic activity or activity in bursts of various length. Thus, for a model, which matches a more realistic PU better, we have to move towards a structured activity model, where the durations of both activity and inactivity of the PU can be described with an appropriate probability distribution.

References

[1] A. Frotzscher, U. Wetzker, M. Bauer, M. Rentschler, M. Beyer, S. Elspass, and H. Klessig, “Requirements and current solutions of wireless communication in industrial automation,” in *IEEE International Conference on Communications Workshops (ICC)*, June 2014.

[2] J. M. Winter, I. Muller, C. E. Pereira, S. Savazzi, L. B. Becker, and J. C. Netto, “Coexistence issues in wireless networks for factory automation,” in *IEEE International Conference on Industrial Informatics (INDIN)*, July 2014.

[3] J. Kim, W. Jeon, K.-J. Park, and J. P. Choi, “Coexistence of full-duplex-based IEEE 802.15.4 and IEEE 802.11,” *IEEE Transactions on Industrial Informatics*, vol. 14, Dec. 2018.

[4] G. Naik, J. Liu, and J.-M. J. Park, “Coexistence of wireless technologies in the 5 GHz bands: A survey of existing solutions and a roadmap for future research,” *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, 2018.

[5] A. Saad, B. Staehle, and R. Knorr, “Predictive medium access control for industrial cognitive radio,” in *IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Jan. 2018.

[6] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor, *Principles of Cognitive Radio*. Cambridge, UK: Cambridge University Press, 2013.

[7] “802.11n-2009 - IEEE standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput,” 2009.

[8] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination function,” *IEEE Journal on Selected Areas in Communications*, vol. 18, Mar. 2000.

[9] C. Song, “Performance analysis of the IEEE 802.11p multichannel MAC protocol in vehicular ad hoc networks,” *Sensors*, vol. 17, Dec. 2017.

[10] W. Jiao, M. Sheng, K.-S. Lui, and Y. Shi, “End-to-end delay distribution analysis for stochastic admission control in multi-hop wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 13, Mar. 2014.

[11] S. Wu, W. Mao, and X. Wang, “Performance study on a CSMA/CA-based MAC protocol for multi-user MIMO wireless LANs,” *IEEE Transactions on Wireless Communications*, vol. 13, June 2014.

[12] K. Sanada, N. Komuro, and H. Sekiya, “End-to-end throughput and delay analysis for IEEE 802.11 string topology multi-hop network using Markov-chain model,” in *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Aug. 2015.

[13] J.-W. Cho and Y. Jiang, “Fundamentals of the backoff process in 802.11: Dichotomy of the aggregation,” *IEEE Transactions on Information Theory*, vol. 61, no. 4, 2015.

[14] S. Geirhofer, L. Tong, and B. M. Sadler, “Cognitive medium access: A protocol for enhancing coexistence in WLAN bands,” in *IEEE Global Telecommunications Conference*, Nov. 2007.

[15] T. Henderson, G. Riley, S. Floyd, and S. Roy, “ns-3 | a discrete-event network simulator for internet systems.” https://www.nsnam.org/.