A Measurement-Based Model for Dynamic Spectrum Access in WLAN Channels

Stefan Geirhofer and Lang Tong
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: {sg355, lt35}@cornell.edu

Brian M. Sadler
Army Research Laboratory
Adelphi, MD 20783-1197
Email: bsadler@arl.army.mil

***Abstract*— In this paper we consider dynamically sharing the spectrum in the time-domain by exploiting whitespace between the bursty transmissions of a set of users, represented by an 802.11b based wireless LAN (WLAN). Realizing that exploiting the under-utilization of the channel requires a good model of the these users' medium access, we propose a continuous-time semi-Markov model that captures the WLAN's behavior yet remains tractable enough to be used for deriving optimal control strategies within a decision-theoretic framework.**

Our model is based on actual measurements in the 2.4 GHz ISM band using a vector signal analyzer to collect complex baseband data. We explore two different sensing strategies to identify spectrum opportunities depending on whether the primary user's transmission scheme is known. The collected data is used to statistically characterize the idle and busy periods of the channel. Furthermore, we show that a continuous-time semi-Markov model is able to capture the data with good accuracy. The Kolmogorov-Smirnov test is used to validate the model and to assess the model's goodness-of-fit quantitatively. A conclusion summarizes the main results of the paper.

I. Introduction

With the proliferation of wireless communication systems, the relevant spectrum has become a scarce resource, most of which has been allocated by regulators as of today. At most times and locations, however, actual measurements show that the spectrum is only lightly used. This apparent paradox gives rise to envisioning different schemes that dynamically reuse the spectrum without causing (significant) interference to the licensee, referred to as the primary user in the following.

In this paper, we consider dynamically sharing the spectrum in the time-domain by exploiting white-space between the bursty transmissions of a primary user [1]. In order not to interfere with the primary user, we need to obtain a reliable model to predict the primary users' access.

The are numerous applications for such a setup. For instance, consider the coexistence of sensor networks with WLAN based applications (or other protocols employing similar MAC schemes). Due to the heterogenous applications of sensor networks and their transient deployment it seems unrealistic to assume that separate frequency bands will be assigned by regulators. However, given that sensor networks usually communicate sporadically and at low rate, it is plausible that the remaining whitespace of other applications could be reused. A similar setup can also be envisioned for different commercial applications such as Bluetooth and WLAN, or to mitigate the effect of interference in unlicensed bands (such as the ISM bands).

[1]Prepared though collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

A. Main contribution

In this work, we propose a model that statistically describes the busy and idle periods of a WLAN. In looking for a model we recognize that there is a fundamental tradeoff between analytical tractability and the accuracy of the model. Ultimately, if we desire to derive an optimal control policy based on such a model (presumably within a decision-theoretic framework), we need to keep the model simple enough. At the same time, however, a tractable model is useless if it is not able to predict the primary user's behavior sufficiently well.

Specifically, we show that a continuous-time semi-Markov process (SMP) represents a good tradeoff between these two extremes. On the one hand extensive research has been performed on the optimal control of semi-Markov decision processes (SMDP) allowing for our model to be embedded in such a framework. On the other hand, the SMP is complex enough to characterize the heavy-tailed distribution of the idle durations appearing in practice. In particular, we show that a generalized Pareto distribution provides for a good fit with the experimental data and we evaluate the goodness-of-fit using the Kolmogorov-Smirnov test.

Finally, we emphasize that our model is based on data obtained from actual measurements with a vector signal analyzer in the 2.4 GHz ISM band. We implement two different sensing strategies and compare their results to confirm the validity of the data. In this way, our approach not only guarantees accurate measurement of the busy and idle periods but also gives valuable insights when it comes to a practical, real-time implementation in hardware.

B. Relevant work

Dynamic spectrum access schemes have recently received considerable attention, both in terms of theoretical concepts as well as practical implementations and testbeds [2]. Sparked by projects such as the DARPA XG program [3] or the European

DRiVE Project [4] different methodologies have been explored to promote efficient use of the spectrum.

Dynamically sharing the spectrum can most efficiently be employed in the spatial or time-domain, each area having its own challenges and limitations. In the spatial domain, the reliable sensing of primary users has to be guaranteed, making it necessary to detect very weak signals [5]. In such a scenario receiver cooperation becomes crucial in many cases [6].

In time-domain applications, the main challenge lies in predicting the medium access of the primary user, and finding optimal control schemes that take practical limitations into account. In [1], [7] optimal control policies are found in a partially-observable decision framework, assuming a Markovian behavior of the primary user.

From a practical viewpoint, numerous testbeds have been set up to explore practical limitations. In [8] a sensing-based DSP/FPGA implementation has been presented for sharing the spectrum with a WLAN.

Statistical modeling in WLAN applications usually focuses on parameters such as packet loss or channel quality, assuming idealized models for the medium access and physical layer. As a result, models proposed within such frameworks do not easily extend to our setup. The class of semi-Markov models considered in this paper, on the other hand, has been employed in numerous applications ranging from characterizing VoIP traffic [9], to speech recognition tasks [10], and related applications.

II. Measurement Setup

In this work we identify the primary user with an 802.11b based WLAN operating in the 2.4 GHz ISM band. Different from existing publications that use commercial WLAN adapter cards to obtain packet traces, we employ a vector signal analyzer (VSA) to capture the raw complex baseband data. Subsequently, we process these data to find the busy and idle periods of the channel. Our approach not only guarantees an accurate and verifiable characterization of the channel but also gives valuable insights into sensing strategies when it comes to a practical implementation.

A schematic of our setup is shown in Fig. 1. The details for both the WLAN setup and the configuration of the VSA are provided in the following sections, respectively.

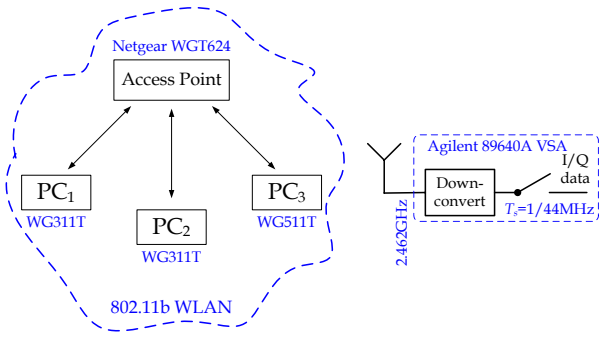


Fig. 1. The measurement setup.

A. WLAN setup

The WLAN consists of a Netgear WGT624 wireless router and three computers with wireless adapter cards (two Netgear WG311T and one WG511T; cf. Fig. 1). The router operated in a 22 MHz frequency band around 2.462 GHz (Channel 11). The router as well as the workstations were located in the same room resulting in a high signal-to-noise ratio (SNR) between nodes and no hidden terminals. Traffic was generated using the Distributed Internet Traffic Generator (D-ITG) [11], which allowed us to statistically characterize parameters such as inter-departure times and packet length.

As we will discuss in detail, our measurements focus on two traffic scenarios. First, we consider high rate UDP traffic from one workstation to the router (the other computers are turned off) to verify the operability of the setup. Subsequently, we consider UDP traffic of constant packet length from all three computers with Poisson distributed inter-departure times at different rates.

B. Vector signal analyzer

In order to capture the transmissions of the WLAN discussed above we use an Agilent 89640A vector signal analyzer to collect complex baseband samples. The device internally downconverts the signal from 2.462 GHz to an internal intermediate frequency (IF) and then samples at a rate of 44 MHz [12]. A large number of continuous blocks of length 0.25 s each are collected in this way and stored for further processing.

III. Sensing Strategies

Given the raw complex baseband data gathered by the VSA, we process the data to find the idle and busy periods of the channel. In particular, we implement two different sensing strategies, depending on whether it is known that the primary user forms a WLAN.

A. Energy-based detection

If the transmission standard of the primary user is not known, a natural approach it to use energy-based detection. To this end we partition the data into blocks of N complex baseband samples and then perform detection for each of them individually (clearly, the blocks must be shorter than the smallest busy/idle duration for this approach to work and for edge-effects to be ignored). We formulate this problem as a hypothesis testing problem between the null-hypothesis of observing only noise V_i and the alternative hypothesis of observing a signal S_i in noise. Mathematically, this is formulated as

$$\mathcal{H}_0 \quad : \quad Y_i = V_i,\ i = 1, \ldots, N \tag{1}$$

$$\mathcal{H}_1 \quad : \quad Y_i = S_i + V_i,\ i = 1, \ldots, N. \tag{2}$$

In our analysis we assume that the noise is drawn from a complex Gaussian distribution with zero mean and variance σ_0^2,

$$V_i \sim \mathcal{CN}(0, \sigma_0^2),\ i = 1, \ldots, N. \tag{3}$$

Considering our assumption that the transmission standard of the primary user is unknown, we further assume that the signal S_i is also drawn from a complex Gaussian distribution with zero-mean but variance σ_1^2,

$$S_i \sim \mathcal{CN}(0, \sigma_1^2), \;\; i = 1, \ldots, N. \tag{4}$$

The above detection problem is standard [13], and the optimal Neyman-Pearson detector is given by

$$T(\mathbf{y}) = \sum_{i=1}^{N} |Y_i|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{5}$$

where the threshold γ is determined by the desired probability of false-alarm. Recognizing that the test statistic $T(\mathbf{y})$ is χ^2 distributed with $2N$ degrees of freedom, we can establish that the probability of false-alarm is given by

$$\alpha = \Pr(T(\mathbf{y}) > \gamma | \mathcal{H}_0) = 1 - \tilde{\Gamma}_r(N, \frac{\gamma}{\sigma_0^2}), \tag{6}$$

where

$$\tilde{\Gamma}_r(N, \xi) = \frac{1}{\Gamma(N)} \int_0^{\xi} t^{N-1} e^{-t} dt \tag{7}$$

is the regularized gamma function and $\Gamma(N)$ is the complete gamma function. Similarly, the power of the detector is given by

$$\beta = \Pr(T(\mathbf{y}) > \gamma | \mathcal{H}_1) = 1 - \tilde{\Gamma}_r(N, \frac{\gamma}{\sigma_0^2 + \sigma_1^2}). \tag{8}$$

The detection performance is influenced by the number of available samples N as well as the relation between σ_0^2 and σ_1^2, which can more conveniently be expressed in terms of the SNR $= \sigma_1^2/\sigma_0^2$. In particular, given desired values for α and β as well as a lower bound on the SNR, we can determine how many samples are needed to arrive at a detector that is accurate enough.

For a minimum SNR of 5 dB, which is a conservative choice for our setup, and blocks of $N = 44$ samples (corresponds to blocks of exactly $1\,\mu$s) we can discern the hypotheses with less than 10^{-5} probability of error. It should be noted however, that the above derivation relies on the idealized assumptions (3) and (4). Given that we have an oversampled signal at hand, especially (4) may require some modification. Additionally, in a practical implementation interference from other channels is an important limitation, especially since the WLAN channels in the 2.4 GHz ISM band are partially overlapping [14]. In practice it can thus be necessary to filter the signal as to mitigate out-of-band interference.

B. Feature-based detection

The last section assumes that we do not have any information on the transmission scheme of the primary user. However, for some applications it is reasonable to assume that we know that we are sharing the spectrum with a WLAN. In this case, we can exploit this knowledge to detect packets more reliably and to extract information on the packet length that is provided in the packet's header.

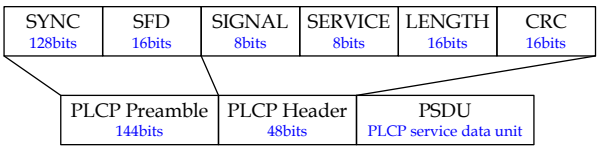


Fig. 2. The PLCP preamble.

The 802.11b standard specifies that a PLCP[1] preamble and header be transmitted with the higher layer frames. The PLCP header contains a synchronization preamble which consists of scrambled '1' bits as well as the start-of-frame delimiter, which indicates the start of the PLCP header. The PLCP header in turn consists of a SIGNAL-field specifying the PSDU's rate of transmission, a SERVICE-field used for standard specific signaling and a LENGTH field which provides the duration in microseconds needed to transmit the payload from higher layers (PSDU) [15]. This LENGTH-field is of particular interest, since it provides us with the exact end of the packet. Together with the SFD this completely determines the start and end of each packet.

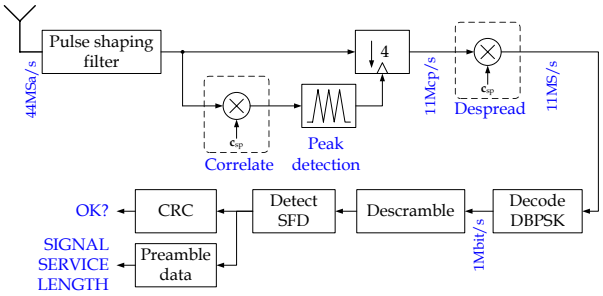
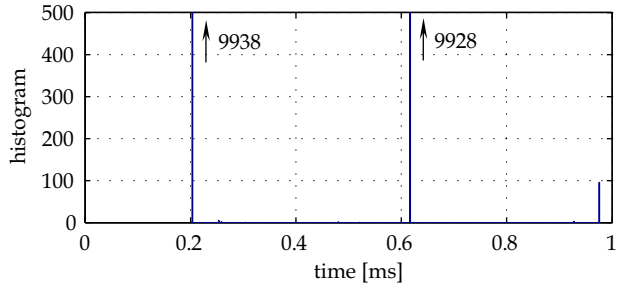


Fig. 3. Receiver chain for the feature-based detection scheme.

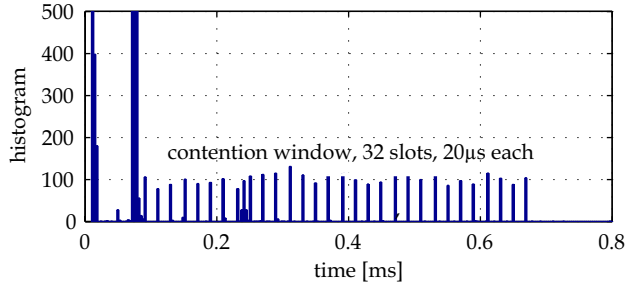
The decoding chain used to decode the WLAN packets is shown in Fig. 3. First, the complex baseband data sampled at a rate of 44 MHz is passed through a Gaussian pulse shaping filter with a bandwidth-symbol time product $BT_s = 1/2$. Subsequently, the signal is correlated with the 11-sample Barker sequence specified by the standard [14]. Consequently, the oversampled signal shows periodic peaks whenever the spreading sequence lines up with the input signal. By detecting these peaks we obtain chip-synchronization. The signal is then downsampled to the symbol rate of 11 Mbps, and despread. While there is a small but noticeable frequency offset between transmitter and receiver, we need not compensate for it given that the preamble is encoded using DBPSK. The decoded bits are then descrambled and the start-of-frame delimiter is detected. Subsequently, the SIGNAL, SERVICE, and LENGTH fields are extracted and the CRC-check is performed to ensure that the extracted information is correct.

It should be noted that we do not claim optimality for the above decoding approach. However, given that we operate at medium to high SNR the above decoding scheme yields good results and mimics the processing used in commercial WLAN

[1]PLCP stands for Physical Layer Convergence Procedure.The router is configured to only use long synchronization preambles.



(a) Histogram for the busy durations.



(b) Histogram for the idle durations.

Fig. 4. Histograms for the measurement validation (cf. Sec. IV-A).

adapter cards.

IV. Measurement results

In this section we present results for the busy and idle periods of the channel obtained by processing the baseband data using the sensing strategies discussed in the last section. First, we focus on an idealized setup with only one terminal and the access point. The results obtained for both the energy and feature-based sensing match very closely, moreover they reflect the characteristics of the WLAN's MAC protocol bolstering the validity of our measurement. Second, we present our measurement results for a setup consisting of three terminals and the access point, for varying data rates.

A. Measurement validation

In order to verify that our measurements reflect the particulars of the WLAN standard, we consider a setup consisting of a single laptop computer with a Netgear WG511T adapter card and the wirless router. The computer generated UDP packets at a rate of 10^5 packets per second with constant inter-departure times and a constant packet length of 512 Bytes. Using the VSA we then collected 40 blocks with a duration of 0.25 s each and analyzed the data using the sensing strategies described above. The durations of idle and busy periods were collected and the histogram of these data is shown in Fig. 4.

First, consider the histogram of the busy periods of the channel in Fig. 4(a). We can see that the packets have three discrete lengths, with approximate durations of 0.2 ms, 0.61 ms, and 0.95 ms, respectively. Indeed this reflects the standardized behavior of the WLAN and the particulars of our setup. In fact, the blocks of length 0.61 ms correspond to the UDP data packets (512 B plus high layer headers), and the blocks of length 0.2 ms correspond to the acknowledgement packets that need to be sent after each successful reception of a data packet. The histograms also shows that there are almost as many acknowledgements as data packets, which indicates that most packets are received successfully. Finally, the significantly smaller component with length 0.95 ms corresponds to periodic beacons that need to be broadcast from the access point for timing purposes. Such packets are sent periodically every 100 ms.

The histogram of the idle periods reflects the particulars of the WLAN standard as well. We see that there are two large, discrete components at $10\ \mu$s and around $70-80\ \mu$s. The former corresponds precisely to the short inter-frame space (SIFS) that separates the transmission of a data packet and its acknowledgement by the receiver. The latter approximately corresponds to the distributed coordination function inter-frame space (DIFS), that is the idle duration that must be waited for before transmitting the next data packet. Finally, we observe 32 discrete components spaced approximately $20\ \mu$s (that is the standardized slot-length) apart. These correspond to the contention window that is used for the medium access of the WLAN.

B. Measurement results

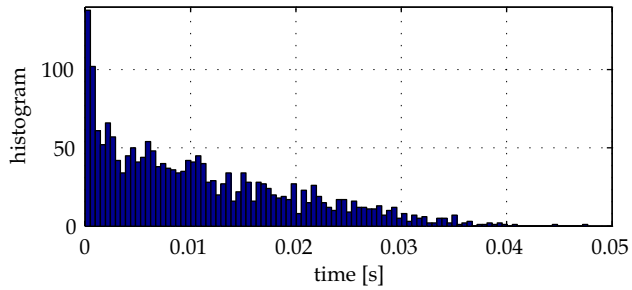
The above section bolsters the validity of our measurement setup and the used sensing strategies. In this section we now use a more representative setup with multiple terminals, each running a traffic generator. While we again used UDP traffic with a constant payload of 512 B, the inter-departure times were chosen to have independent Poisson distributions with varying rate parameter λ. For the busy intervals we again observe a very similar histogram compared to Fig. 4. However, the idle times are now highly dependent on the parameter λ. Ultimately, as the traffic rate increases the channel becomes busier and the available whitespace decreases.

This behavior is illustrated in Fig. 5, where the upper and lower figure correspond to $\lambda = 25$pkts/s and $\lambda = 500$pkts/s, respectively. We can observe that in the low rate case, the histogram suggests a heavy-tailed behavior; we will address this fact in more detail in the next section. On the other hand, at high rate, the histogram more resembles an exponential distribution.

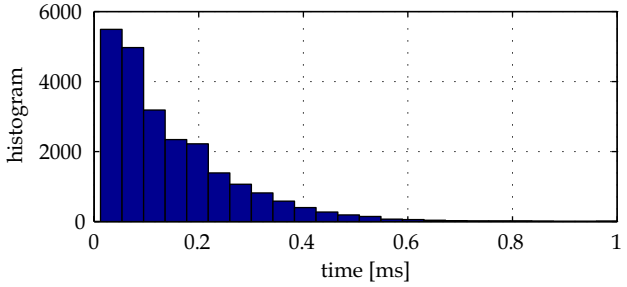
V. Proposed semi-Markov model

The measurement results presented in the last section allow for two important conclusions when it comes to establishing a model. First, we can categorize the state of the channel by four different states. In particular we assume that the channel can be busy due to either the transmission of a data packet, or due to the transmission of an acknowledgement. Second, the channel is either idle for just the short inter-frame space, or for an extended period either corresponding to stations contending for the medium, or the fact that none of the terminals has data to transmit. The state space of the model is illustrated in Fig. 6.

From the viewpoint of a secondary user, it is especially important to characterize how long the channel remains in each



(a) Idle periods for $\lambda = 25$pkts/s



(b) Idle periods for $\lambda = 500$pkts/s

Fig. 5. Histograms of the idle durations for small and large λ.

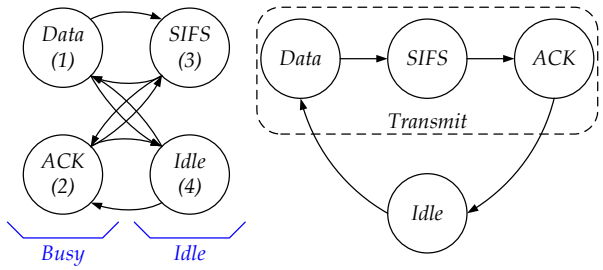


Fig. 6. Proposed Markov model. The lumped model (with deterministic DATA→SIFS→ACK transitions is shown on the right).

state. Given that the idle times show a continuous behavior, a continuous-time Markov process (CTMP) appears to be an intuitive approach. However, considering the heavy-tailed behavior of the idle periods (cf. Fig. 5), we can infer that this class of model is not appropriate since a CTMP requires the sojourn times in each state to be exponentially distributed [16]. Consequently, we will consider a continuous-time semi-Markov process, which allows for an arbitrary specification of the sojourn time distribution in each state.

More precisely, a semi-Markov process is a stochastic process whose transition behavior can be characterized in two steps [17]. First, the transition between states follow a Markov chain and are hence specified by a transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}, \tag{9}$$

where p_{ij} denotes the probability that a transition from state i to state j occurs. Secondly, given that the system is in state i and will transition to state j, the sojourn time t in state i is distributed according to cumulative distribution function $Q_{ij}(t)$. Note that $p_{ii} = 0$ for all i because the arbitrary specification of the sojourn time fully captures the time spent in each state.

A. Estimating transition probabilities

The first step in characterizing the channel as a semi-Markov process is to estimate the transition probabilities from our measurement data. Given the idle and busy durations we first classify every observation according to the four states of our model. Then, given this unbroken chain of observations, we can use well-known maximum-likelihood techniques [18] to obtain estimates for the transition probability matrix.

In particular, consider the following estimator

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \tag{10}$$

where the *transition count* n_{ij} is the number of transitions $i \to j$ occuring in our observation sequence. Similarly, $n_i = \sum_k n_{ik}$ is the number of times that the system resides in state i.

Using the above equation, we can estimate the transition matrix given our observation sequence. The transition matrix is essentially constant with respect to the traffic rate λ. In the case of $\lambda = 100$pkts/s we have

$$P = \begin{bmatrix} 0 & 0 & 0.960 & 0.040 \\ 0 & 0 & 0.003 & 0.997 \\ 0.002 & 0.998 & 0 & 0 \\ 0.996 & 0.004 & 0 & 0 \end{bmatrix} \begin{matrix} \xleftarrow{\text{Data}} \\ \xleftarrow{\text{ACK}} \\ \xleftarrow{\text{SIFS}} \\ \xleftarrow{\text{Idle}} \end{matrix}. \tag{11}$$

We can see that for our measurement setup with high SNR transmission between nodes and no hidden terminals the sequence of states DATA→SIFS→ACK is essentially deterministic (the corresponding transition probabilities are very close to one). Hence, it is possible to simplify the model by lumping these states together. While this inhibits us to model the occurrence of collisions, we retain good accuracy since collisions are infrequent.

According to the above we consider the simplified model shown in Fig. 6, which consists of a 'transmit' state (a lumped version of DATA, SIFS, and ACK states with deterministic transitions), and an idle state. The transition probabilities for this simplified semi-Markov model are now trivial, since every transmit state *must* be followed by an idle period. Consequently, to fully specify the semi-Markov model, we only need to characterize the sojourn times in each state.

B. Fitting sojourn distributions

Given the simplified model shown in Fig. 6 it remains to specify the distribution $Q_{ij}(t)$ for the sojourn times in each state. This is easily done for the TRANSMIT state since the sojourn time in the DATA, SIFS, and ACK state are all deterministic, we also have a deterministic sojourn time in the TRANSMIT state.

The sojourn time in the IDLE state is more difficult to approximate. The histograms in Fig. 5 show that for small λ

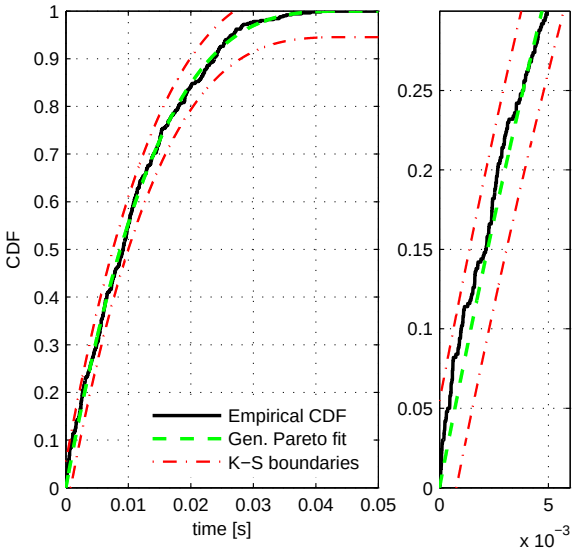


Fig. 7. Fitting the sojourn time in the IDLE-state for $\lambda = 25$pkts/s. The right figure shows an enlarged view for small t. The K-S boundaries are drawn for $p = 0.1$.

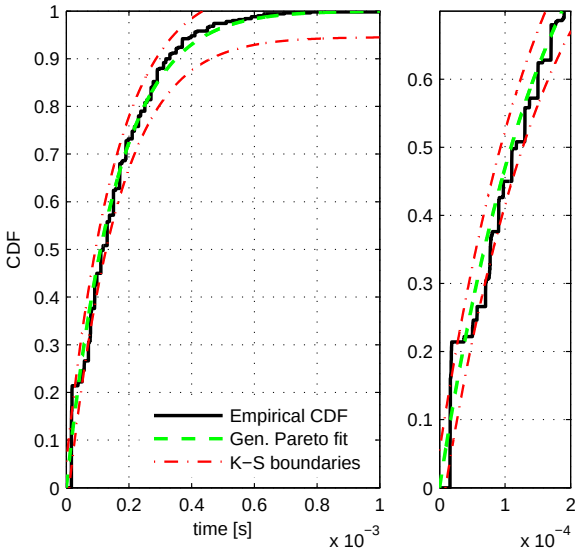


Fig. 8. Fitting the sojourn time in the IDLE-state for $\lambda = 500$pkts/s. The right figure shows an enlarged view for small t. The K-S boundaries are drawn for $p = 0.1$.

we see heavy-tailed behavior, while for large λ an exponential distribution seems to be a good fit. Since we are concerned with finding a distribution that provides an overall good fit, we focus on distributions whose shape can approximate both extremes. In particular, we consider a generalized Pareto distribution with probability density function [19]

$$f(t|k,\sigma) = \frac{1}{\sigma}\left(1 + k\frac{t}{\sigma}\right)^{-1-1/k}, \tag{12}$$

where $k \neq 0$ denotes the shape parameter, and σ represent the scale parameter. We assume that the threshold parameter θ is zero. It should be noted that for $k = 0$ the generalized Pareto distribution converges to the exponential distribution.

The parameters for the generalized Pareto distribution are estimated using maximum-likelihood techniques. The parameter estimates are shown in Tab. I. The fitted cumulative density function are shown in Fig. 7 and Fig. 8 for $\lambda = 25$ pkts/s and $\lambda = 500$ pkts/s, respectively. The generalized Pareto distribution shows a good fit in both cases. For comparison, an exponential distribution is also fit to the data. While for large λ the goodness-of-fit is comparable to the Pareto distribution, an exponential assumption is clearly inappropriate for small λ.

C. Goodness-of-fit test

In order to validate the goodness-of-fit of the distributions fit to the empirical data in the last section, we consider the Kolmogorov-Smirnov (K-S) test. Given n independent random variables $Y_1, \ldots, Y_n$, this test is a well-known technique for discerning the hypotheses

$$\mathcal{H}_0 \quad : \quad Y_i \sim F(t),\ i = 1, \ldots, n \tag{13}$$

$$\mathcal{H}_1 \quad : \quad Y_i \nsim F(t),\ i = 1, \ldots, n, \tag{14}$$

that is determining whether y_i are drawn from a given continuous distribution function F. Specifically, given the observations we first construct the empirical distribution function

$$F_e(t) = \frac{\#i : y_i \leq t}{n}, \tag{15}$$

where $\#i : y_i \leq t$ denotes the number of observations smaller than t. If the observations are indeed drawn from distribution F, we would assume that $F_e(t)$ and $F(t)$ match well. To assess the deviation from this hypothesis consider the K-S test statistic

$$D = \max_t |F_e(t) - F(t)|. \tag{16}$$

Naturally, a small value of D suggests that the samples are indeed drawn from distribution F. While the quantity D reflects the goodness-of-fit, we usually consider the so-called p-value, which is defined as

$$p = \Pr(D \geq d|\mathcal{H}_0). \tag{17}$$

It can be shown that (17) is independent of the distribution F and, given $D = d$ can easily be evaluated by simulation [20]. Usually, a value of $p \approx 0.1$ is deemed high enough as to assume that the observations are indeed drawn form distribution F.

We applied the K-S test described above to our problem identifying $F(t)$ with the fitted distribution and $F_e(t)$ as the empirical distribution function found by measurement. The obtained values for D and p are also summarized in Tab. I.

While from Fig. 7 and Fig. 8 the generalized Pareto distribution seems to be a good fit, the p-values obtained through the K-S test seem rather small and do not reach the typical value of 0.1. Nevertheless, this discrepancy can be attributed to the presence of the contention window. In fact, the idle periods of the channel (disregarding the SIFS) will be a mixture of the idle times due to the contention window and those that

λ [pkts/s]	25	50	75	150	300	500
Generalized Pareto fit						
k	-0.3018	-0.3123	-0.3095	-0.2662	-0.0239	-0.0455
σ	0.0139	0.0072	0.0045	0.0017	$2.35 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$
D	0.0368	0.0844	0.0572	0.0798	0.1180	0.0961
p-value	0.4958	0.0015	0.0728	0.0032	0	0
D^*	0.0368	0.0683	0.0388	0.0396	0.0550	0.0281
p-value*	0.4956	0.0180	0.4281	0.4029	0.0936	0.8138
Exponential fit						
k	0.0107	0.0054	0.0034	0.0013	$2.30 \cdot 10^{-4}$	$1.53 \cdot 10^{-4}$
D	0.0813	0.0601	0.0797	0.0519	0.1226	0.1011
p-value	0.0025	0.0518	0.0033	0.1307	0	0
D^*	0.0813	0.0600	0.0794	0.0515	0.0536	0.0283
p-value*	0.0025	0.0525	0.0035	0.1361	0.1091	0.8070

TABLE I

Parameter estimates for the fitted generalized Pareto and exponential distributions, respectively. For both cases, the K-S statistic D and corresponding p-values are shown.

are really due to the fact that none of the terminals has data to transmit (we shall refer to the second scenario as a 'free' channel for brevity).

The fact that the distribution of the idle periods really is a mixture of two different distributions has different effects on the CDF, for large and small λ, respectively. At small λ, the effect of the contention window is visible as an increased slope at small values of t, thus affecting the D value obtained by the K-S test. A more appropriate goodness-of-fit measure is thus to consider only the deviation for idle times that are larger than the contention window, i.e.

$$D^* = \max_{t \geq \tau} |F_e(t) - F(t)|. \tag{18}$$

If we choose $\tau = 2 \cdot 10^{-3}$ for $\lambda \leq 150$, we obtain alternative values of for the K-S statistics, denoted by D^* in Tab. I.

For increasing λ, however, the contention window becomes the dominant component (cf. Fig. 8). While in this case the generalized Pareto distribution still nicely approximates the empirical CDF, we see that this function is not smooth but resembles a 'staircase'. This can be attributed to the slot structure of the contention window with discrete components space 20μs apart. Consequently, we really are approximating a discrete probability distribution with a continuous one and have to expect some residual error. In fact, given the steep slope of the CDF, the first few steps largely influence the resulting D-value (cf. Fig. 8). By disregarding these components according to (18), we can again find modified values D^* that provide for a more appropriate evaluation.

Given the values for D^*, we can again evaluate (17) to find the modified p^*-values, again shown in Tab. I. Based on the p^*-values the generalized Pareto distribution appears to be an appropriate fit for all λ's. In contrast, the exponential distribution is a poor fit fit for small λ.

VI. Conclusion

In conclusion, we have proposed a continuous-time semi-Markov model that captures the idle periods remaining between the bursty transmissions of a wireless LAN. We demonstrate that a generalized Pareto distribution provides for an adequate fit for varying packet rates λ.

Furthermore, we believe that the model strikes a good compromise between accuracy and computation complexity. In particular, our model can be applied within the framework of semi-Markov decision processes, allowing for deriving optimal control polices.

In the future we plan to evaluate the model's goodness-of-fit for more realistic traffic scenarios, including HTTP and FTP-traffic, as well as streaming applications (Voice-over-IP, video conferencing, etc.).

References

[1] Q. Zhao, L. Tong, and A. Swami, "Decentralized Cognitive MAC for Dynamic Spectrum Access," in *Proc. First IEEE Internatial Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 224–232.

[2] *Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005.

[3] "DARPA: The Next Generation (XG) Program," http://www.darpa.mil/ato/programs/xg/index.htm.

[4] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, "DRiVE-ing to the Internet: Dynamic Radio for IP Services in Vehicular Environments," in *Proc. of the 25th Annual IEEE Conference on Local Computer Networks*, Nov. 2000, pp. 281–289.

[5] R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *Proc. WirelessCom 05 Symposium on Signal Processing*, 2005.

[6] A. Sahai, N. Hoven, S. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse," *submitted to IEEE J. Select. Areas Commun.*, 2006.

[7] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," *submitted to IEEE J. Select. Areas Commun.*, Feb. 2006.

[8] S. D. Jones, N. Merheb, and I.-J. Wang, "An experiment for sensing-based opportunistic spectrum access in CSMA/CA networks," in *First IEEE Internatial Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 593–596.

[9] A. Biernacki, "VoIP Source Model based on the Hyperexponential Distribution," *Enformatika, Trans. on Engineering, Computing, and Technology*, vol. 11, pp. 202–206, Feb. 2006.

[10] S. E. Levinson, "Continuously variable duration hidden Markov models for speech analysis," in *Proc. International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, Apr. 1986, pp. 1241–1244.

[11] S. Avallone, A. Botta, D. Emma, S. Guadagno, and A. Pescape, "D-ITG V.2.4 Manual," University of Napoli "Federio II", Tech. Rep., Dec. 2004.

[12] Agilent Technologies, "Agilent 89611A 70MHz IF Vector Signal Analyzer," data sheet, Oct. 2001.

[13] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, 1994.

[14] ANSI/IEEE Standard 802.11b-1999 (R2003), "Wireless lan medium access control (mac) and physical layer (phy) specifications: Higher-speed physical layer extension in the 2.4ghz band," IEEE SA Standards Board, Tech. Rep., 1999.

[15] ANSI/IEEE Standard 802.11, 1999 Edition (R2003), "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE/SA Standards Board, Tech. Rep., 1999.

[16] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw Hill Publishing Company, 2002.

[17] S. M. Ross, *Applied Probability Models with Optimization Applications*. Dover Publications, 1970.

[18] P. Billingsley, "Statistical Methods in Markov Chains," *The Annals of Mathematical Statistics*, vol. 32, no. 1, pp. 12–40, Mar. 1961.

[19] S. Kotz and S. Nadarajah, *Extreme Value Distributions. Theory and Applications*. Imperial College Press, 2000.

[20] S. M. Ross, *Simulation*, 3rd ed. Academic Press, 2002.

[2]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.