Ad Hoc Networks 21 (2014) 60–83

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Closing the gap between traffic workload and channel occupancy models for 802.11 networks

I. Glaropoulos [a,*], A. Vizcaino Luna [a], V. Fodor [a], M. Papadopouli [a,b]

[a] School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
[b] Department of Computer Science, University of Crete, Heraklion, Greece

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form 17 March 2014
Accepted 4 May 2014
Available online 14 May 2014

Keywords:
802.11 Networks
Traffic models
Channel occupancy models
Statistical validation

ABSTRACT

The modeling of wireless network traffic is necessary to evaluate the possible gains of spectrum sharing and to support the design of new cognitive protocols that can use spectrum efficiently in network environments where diverse technologies coexist. In this paper we focus on IEEE 802.11 wireless local area networks and close the gap between two popular levels of modeling, macroscopic traffic workload modeling and microscopic channel occupancy modeling. We consider traffic streams generated by established traffic workload models and characterize the networking scenarios where a simple, semi-Markovian channel occupancy model accurately predicts the wireless channel usage. Our results demonstrate that the proposed channel occupancy model can capture the channel idle time distribution in most of the scenarios, while the Markovian assumption cannot be validated in all cases.

1. Introduction

Spectrum sharing among diverse network technologies has been introduced as a promising solution to increase the efficiency of spectrum utilization in wireless environments, and thus ease the problem of spectrum scarcity. One of the key components of efficient spectrum sharing is cognitive medium access control, building on the knowledge of the channel usage patterns of the coexisting networks [1]. Therefore, traffic workload and channel occupancy models, either considered to be known [2] or derived on-line [3], are necessary for protocol design and channel access optimization. The issue of network coexistence in the open ISM band is particularly relevant due to the proliferation of diverse low-power wireless technologies, all sharing the ISM spectrum with the high-power Wireless Local Area Networks (WLANs). Accurate WLAN modeling enables these low-power technologies to alleviate harmful WLAN interference [4] and to ensure an effective use of the shared open spectrum [5,6].

WLAN modeling can be classified in two main categories, based on the considered time scale, *traffic workload* modeling and *channel occupancy* modeling. *Traffic workload* studies involve stochastic analysis and modeling of high-layer traffic statistics, such as user arrival and departure process [9] and client-generated flow statistics [10–12], the characterization of user traffic [8,13] or the user mobility [14,15]. In these studies WLAN measurement data is collected via active probing or passive network monitoring, followed by the statistical processing of the collected data, when analytic probability distributions are fitted to the empirical traces. Traffic workload models are often specific to a given networking scenario, for example [16] considers a campus-wide WLAN and provides detailed multi-level, campus-wide WLAN traffic modeling where both session

* Corresponding author. Tel.: +46 76 232 2543.
E-mail addresses: ioannisg@kth.se (I. Glaropoulos), alvl@kth.se (A. Vizcaino Luna), vfodor@kth.se (V. Fodor), mgp@ics.forth.gr (M. Papadopouli).

http://dx.doi.org/10.1016/j.adhoc.2014.05.001
1570-8705/

and flow statistics are collected and fitted to analytic distributions. Although sufficiently realistic, these approaches capture the behavior of WLANs only at a macroscopic level.

Contrary to traffic workload modeling, *channel occupancy* studies aim at modeling directly the short term temporal behavior of the channel status in WLAN networks. They characterize the periods when the channel is either *active* due to a WLAN packet transmission, or *idle*. Clearly, the distribution of the active times is determined by the packet sizes used by the applications, while the distribution of the idle periods depends on both the process of packet generation and the medium access control. We can distinguish between *analytic* and *measurement-based* studies, depending on whether the spectrum occupancy model is developed based on analytic modeling of user behavior and network protocols, or it is extracted from channel occupancy measurements. The seminal work in [17] gives an analytic model for the impact of the IEEE 802.11 MAC protocol on channel occupancy, and derives the network throughput of a single Access Point (AP) WLAN assuming saturated user traffic, i.e. users always have packets to transmit. The case of a non-saturated single WLAN AP is studied in [18], modeling the packet arrivals at the users as a Bernoulli process. In [19] WLAN output buffers are modeled as M/G/1 queues, resulting in sub-geometric idle period distribution. The generality of these analytic channel occupancy models, is, however, limited, since they are based on specific, simple traffic workload models.

As far as measurement-based approaches are concerned, in [20] a hyper–exponential distribution is fitted to the empirical idle period distribution derived by traffic traces from an area with heterogeneous wireless devices. In [21] a Markovian channel occupancy model is developed based on channel measurements extracted from controlled laboratory environments in the 2.4 GHz ISM band. In [22,23] the heavy-tailed behavior of the idle channel periods is demonstrated and a mixture distribution is proposed to capture the two basic sources of channel inactivity, the short, almost uniformly distributed contention windows and the long, heavy-tailed white space periods, when the WLAN users are inactive. The simplicity of the resulting semi-Markovian model makes it attractive for analytic performance studies and cognitive protocol design [5,23,24]. This model considers an idle period length distribution with a high number of degrees of freedom and, potentially, good fitting quality, therefore we select it as the candidate channel occupancy model. In [22,23] it has been validated for a limited set of scenarios, under perfect channel conditions and considering constant UDP payload traffic with exponential packet inter-arrival times. In this paper we evaluate it for a wide range of traffic patterns and network scenarios and define the key factors that affect its accuracy. As traffic traces cannot provide the diversity we are looking for, we build our evaluation on synthetic traces based on validated traffic workload models. To investigate the generality of the model, we select three scenarios with significantly different traffic workload characteristics, namely, university campus, conference-hall, and industrial-plant WLANs.

Specifically, we validate the proposed heavy-tail idle period distribution and the Markovian *assumption*, that is, the assumption that consecutive idle periods have independent durations. We focus on the idle periods, as the active period distribution is not significantly affected by the medium access control, instead, it is directly determined by the packet sizes in the application mix. The model validation is completed with an evaluation of its accuracy considering a restricted dataset of real WLAN traces.

The main contributions of the paper are summarized as follows:

(1) *Traffic workload and channel occupancy modeling.* We define detailed traffic workload models for the three networking scenarios, and parameterize the related semi-Markovian channel occupancy model using extensive simulations of an IEEE 802.11 AP.
(2) *Evaluation of the idle period distribution.* We evaluate the fitting quality of the channel occupancy model, specifically considering the distribution of the idle times. Our results indicate that the mixture distribution proposed in [22,23] is valid in a wide range of networking scenarios.
(3) *Evaluation of the Markovian assumption.* We evaluate the validity of the Markovian assumption considering the correlation of the consecutive idle time period lengths. We conclude that the idle period lengths may be correlated at low or very high load and when the traffic is highly heterogeneous. Therefore the Markovian assumption has to be applied with care.
(4) *Model validation with real WLAN traces.* We evaluate the fitting accuracy of the semi-Markovian model as well as the validity of the Markovian assumption, considering a set of real 802.11 channel occupancy traces captured in diverse WLAN environments. The results are similar to the ones with the synthetic traces, however, they show as well that real occupancy traces can reflect unexpected traffic characteristics.

The remainder of the paper is structured as follows. In Section 2 we review the considered traffic workload and channel occupancy models. In Section 3 we introduce the networking scenarios under study, along with a detailed description of the multi-layer traffic models. The simulation setup, as well as the employed statistical validation tools are presented in Section 4. Section 5 presents the results of the channel occupancy model validation using synthetic WLAN channel occupancy traces, while Section 6 includes the validation over the real WLAN traceset. Section 7 concludes the paper.

2. Traffic workload and channel occupancy models

In this section we define the structure of the multi-layer WLAN traffic workload model and the analytic model for WLAN channel occupancy that is considered in the paper.

2.1. Multi-layer WLAN traffic workload model

Fig. 1 depicts the structure of the multi-layer traffic workload model. As suggested by [16], the *sessions*

(Fig. 1(a)) on the top of the model hierarchy represent the users of the WLAN. Sessions *arrive* at the network following a stochastic process that is, in general, time-variant. A session is characterized by the *number* of traffic *flows* it generates (Fig. 1(b)), and by the *inter-arrival times* between these flows. The length of the sessions is not part of the model. A session is considered to be terminated, when its last flow ends.

A flow is defined as the unidirectional end-to-end packet sequence from a specific transport-layer source–destination connection or a media stream. Each flow is characterized by its *size* in bytes (Fig. 1(b)), and the *in-flow characteristics* given by the sizes, and the inter-arrival times of the *packets* (Fig. 1(c)). The in-flow characteristics depend on the particular network application that generates the flow [13]. Note that the effect of the medium access control is not considered in the traffic workload model.

2.2. Semi-Markovian channel occupancy models

Fig. 2 depicts the IEEE 802.11 WLAN channel occupancy model, originally proposed in [22,23]. The *states* of the semi-Markovian system correspond to the different phases of the WLAN transmission cycle. The channel is *Active* when there is a packet, either Data or ACK, under transmission. Neglecting the short Inter-frame Space (*SIFS*) period between a Data and an ACK packet, due to its considerably short duration, *Data*, *SIFS* and *ACK* states can be merged together into a single *Active* state. In the absence of packet transmission the WLAN channel is *Idle*, and we distinguish between the short *back-off* idle periods (*BK*) introduced by the IEEE 802.11 contention resolution mechanism, and the significantly longer idle periods due to user inactivity, denoted as WLAN *white spaces* (*WS*). Merging the states *BK* and *WS* into a single *Idle* state, the model is reduced into a two-state, semi-Markovian model with holding times $f_A(t), f_I(t)$, respectively. As proposed in [22] active periods are sufficiently modeled as uniform, while $f_I(t)$ needs to be described by a *mixture* distribution, which aims at capturing both of the sources of channel inactivity, that is, the back-off periods with holding times $f_I^{(\mathrm{BK})}(t)$ and the white space periods with $f_I^{(\mathrm{WS})}(t)$. Consequently, $f_I(t)$ obtains the form:

$$f_I(t) = pf_I^{(\mathrm{BK})}(t) + (1-p)f_I^{(\mathrm{WS})}(t), \quad t \geqslant 0, \tag{1}$$

with the *mixture* parameter $p \in [0, 1]$ defining the probability that an idle period is a short back-off.

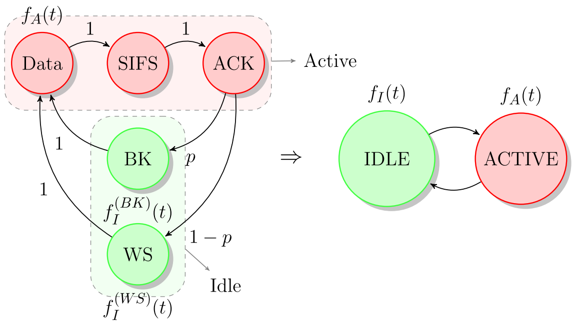


Fig. 2. The semi-Markovian channel occupancy model and its two-state representation.

Measurement results in [22] suggest that the back-off period durations can be modeled with uniform distribution:

$$f_I^{(\mathrm{BK})}(t) = 1/\alpha_{\mathrm{BK}}, \quad \alpha_{\mathrm{BK}} = 0.07 \text{ ms}, \tag{2}$$

independently from the network load. Note that α_{BK} is not the maximum possible back-off period in IEEE 802.11 WLANs, however, higher values appeared rarely in the measurement set [22]. The white space periods are suggested to be modeled by a zero-location, generalized Pareto distribution:

$$f_{\mathrm{gP}}(t) = \frac{1}{\sigma}\left(1 + \xi\frac{t}{\sigma}\right)^{-\frac{1}{\xi}-1}, \tag{3}$$

with σ and ξ, being the *scale* and *shape* parameters, that depend on the actual network traffic. To capture the effect of access-point *beaconing*, in this paper we consider a truncated version of (3). As white spaces are limited by the beacon *period*, T_B:

$$f_I^{(\mathrm{WS})}(t) = \frac{1}{\sigma F_{\mathrm{gP}}(T_B)}\left(1 + \xi\frac{t}{\sigma}\right)^{-\frac{1}{\xi}-1}, \quad t \in (0, T_B), \tag{4}$$

where $F_{\mathrm{gP}}(T_B) \triangleq Pr[T_B \leqslant t] = \left[1 - \left(1 + \frac{\xi}{\sigma}t\right)\right]^{-1/\xi}$. The average white space duration is:

$$\bar{I}^{(\mathrm{WS})} = \int_0^{T_B} f_I^{(\mathrm{WS})}(t)dt \approx \frac{\sigma}{1-\xi}, \quad \text{if } F_{\mathrm{gP}}(T_B) \approx 1. \tag{5}$$

Note, that this channel occupancy model considers good channel conditions without the eventual loss of Data or ACK packets. Under packet losses, the Active and Idle period length distributions are slightly different. Specifically, first, the Active period may not include the ACK packets,

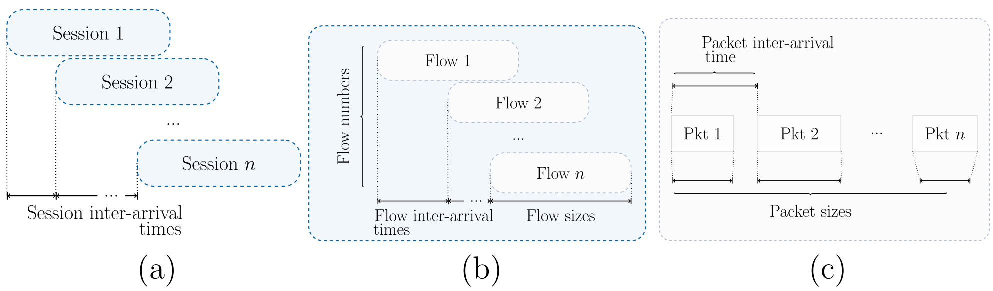


Fig. 1. The structure of the multi-layer traffic workload model, comprising of (a) session-, (b) flow- and (c) in-flow processes.

and therefore, the Active time distribution is affected by the packet loss probability. Note, however, that the estimation is still straightforward, since the distribution depends only on the Data packet sizes in the original traffic mix and on the packet loss probability. Still, the length of the ACK packets is very short compared to the size of the majority of the Data packets. Second, in the case of Data packet loss ACKs are not transmitted, and the idle period starts with a short idle time determined by the Net Allocation Vector (NAV) timer, typically set as the SIFS + ACK duration, in the order of tens of microseconds, followed by the minimum back-off period. Consequently the maximum of the back-off period (α_{BK}) is not affected. The loss affects p, the ratio of back-off periods, but this is a parameter to be estimated. Third, if the ACK is transmitted, but not received correctly, only p is affected due to the packet retransmission. Consequently, the proposed model fits well also for modeling the idle-time distribution in scenarios with transmission errors.

2.3. Validation of the channel occupancy model

We assess the ability of the aforementioned semi-Markovian channel occupancy model to capture the statistical behavior of a practical WLAN channel usage, focusing on the idle time distribution, $f_I(t)$. This assessment involves the following steps:

(1) We generate synthetic traffic streams based on the traffic workload model in Section 2.1, simulate the stream transmissions on a IEEE 802.11-compliant WLAN, and extract the resulting idle time periods sequence.
(2) Then, using the collected idle period samples, we determine T_B as the maximum idle period sample,and estimate the parameters of $f_I(t)$, that is ξ and σ, the *shape* and *scale* parameters of the generalized Pareto distribution, and the *mixture* p [25].
(3) We preform a detailed statistical analysis over the simulated and analytic idle period traces to evaluate the accuracy of the channel occupancy model.

3. Networking scenarios

In this paper we focus on three common WLAN networking scenarios, where the traffic workload is expected to be different:

(1) campus WLAN,
(2) conference-hall WLAN,
(3) industrial-plant WLAN.

3.1. Campus WLAN

To model the traffic demand in the campus WLAN, we follow the session arrival and flow models of [16]. Table 1 summarizes the proposed distributions and their parameters. As [16] shows the session arrival intensity varies during the day. However, it does not change significantly within the time frame at which traffic modeling is useful for protocol design and evaluation. Therefore the session arrivals are modeled by a stationary Poisson process with parameter value drawn from a truncated geometric distribution to model arrival intensities reported in [16]. Flow inter-arrival times are log–normal, while the number of flows per session and the flow sizes are modeled by heavy-tailed BiPareto distributions.

As [16] does not discuss in-flow characteristics, we categorize the flows according to [13] as DNS, Web, FTP, P2P, VoIP, and Video, which gives a highly heterogeneous traffic mix, and give the probability that a flow belongs to a given category. We model the DNS flow as the transmission of a single, uniform-sized packet, since DNS requests and replies typically have short payload. We characterize the Web flows with uniform packet sizes and exponential packet inter-arrival times, accounting for persistent HTTP connections. FTP flows consist of almost deterministic-size packets, transmitted back-to-back. To characterize P2P flows, we follow the suggestions in [26] assuming BitTorrent traffic. Finally, we model the VoIP and Video stream flows considering that they represent Skype traffic and follow the suggestions in [27] on Skype video and voice traffic flow characteristics. The parameters are summarized in Table 2.

3.2. Conference-hall WLAN

While several conference-hall traffic workload models have been proposed, they do not follow the structure proposed in [16]. Therefore, we consider the conference-hall WLAN model proposed in [7], and convert it to the structure proposed in [16]. The conference-hall model suggests that the session arrival process follows the conference time schedule, and proposes statistical distributions for the characterization of the *session arrival process*, the *session durations*, and the *data rate of a session*. It also categorizes the flows according to applications, and gives the probability distribution of the *application mix*. Table 3 summarizes the modeling of the traffic workload. Session arrivals are modeled by an ON/OFF Markov-Modulated Poisson Process (MMPP), where the ON state represents the beginning of a conference event. The session durations are Pareto distributed, while the session data rates are randomly selected within three different rate intervals, representing *light*, *medium* and *heavy* traffic, with probabilities p_L, pM, and p_H, respectively. Once the rate interval is chosen, the data rate is uniformly selected inside the given interval. The application mix is characterized by the probability that a flow belongs to a given application category and additionally by the portion of the traffic an application category generates.

While this model includes the notion of a traffic flow through the definition of the application mix, it does not give models for the number of flows per session, for the flow size and for the flow arrival process. We, therefore, convert the model to the multi-layer traffic workload model by characterizing the missing model elements as in [16] and by parameterizing their statistical distributions in order to fit with the given session, transmission rate and application mix characterization. For each session we, first, draw the session duration and the data rate according to the model parameters proposed in [7], and calculate the

Table 1
Campus WLAN. Traffic workload model parameters.

Traffic object	Distribution	Parameters
Session		
Session arrival (h)	Stationary Poisson process (λ)	1 (min), 928 (max), 11 (median) trunc. Geometric
Flow		
Flow inter-arrival times (s)	Log–normal	$\mu = -1.6355, \sigma = 2.6286$
Flow numbers per session	BiPareto	$\alpha = 0.07, \beta = 1.75, c = 295.38, k = 1$
Flow sizes (bytes)	BiPareto	$\alpha = 0.00, \beta = 1.02, c = 15.56, k = 111$
Application		
	Web/FTP/P2P VoIP + Video/Other	$PWeb = 0.25, P_{FTP} = 0.12,$ $P_{P2P} = 0.33, P_{V+V} = 0.21,$ $P_{oth} = 0.09$

Table 2
Packet size and inter-arrival time distributions for the various network applications.

Application	Packet size (bytes)	Inter-arrival times (s)
DNS	Deterministic (512)	Single packet
Web	Uniform (512, 1536)	Exponential ($\lambda = 10^{-1}$)
FTP	Uniform (892, 1152)	Deterministic (10^{-3})
P2P	Exponential ($\lambda = 512$)	Weibull (0.53, 0.13532)
VoIP	Uniform (128, 384)	MMDP (λ, μ, T) $\lambda = 0.1, \mu = 0.1, T = 0.03$
Video	Uniform (384, 768)	MMUP ($\lambda, \mu, \alpha, \beta$) $\lambda = 0.1, \mu = 0.1, \alpha = 0.03, \beta = 0.05$

Table 3
Conference-hall WLAN. Traffic workload model parameters.

Traffic object	Distribution	Parameters
Session		
Session arrival (s^{-1})	ON/OFF MMPP [$\lambda_{ON}, \lambda_{OFF}$]	$\lambda_{ON} = 38^{-1}$, $\lambda_{OFF} = 0$
Session duration (min)	Pareto	$\xi = 0.78, \sigma = 30.76$
Session workload		
Data rate (kbps)	Uniform *light* (p_L = 0.25), *medium* ($p_M = 0.65$), *heavy* ($p_H = 0.1$)	 $\mu_l = 15, \max_l = 60$ $\mu_m \in (15 - 80)$, $\max_m \in (60 - 175)$, $\max_h = 175, \mu_h = 80$
Application		
	Web$_{(HTTP,HTTPS,SSH)}$/ DNS$_{(ICP,ICMP,DNS)}$	$P^{fl}_{Web} = 0.62, P^{fl}_{DNS} = 0.31$ $P^{fl}_{oth} = 0.07$ $P^{vol}_{Web} = 0.68, P^{vol}_{DNS} = 0.11$ $P^{vol}_{oth} = 0.21$

traffic volume (in number of bits) of the session. Given the traffic volume and the predefined application mix, we first determine the number of DNS flows, based on our DNS model in Table 2. Then, from the number of DNS flows we estimate the number of Web and other flows and the respective flow sizes, to fit the application mix parameters. Finally we parameterize the flow inter-arrival time distribution to match the known session duration.

3.3. Industrial-plant WLAN

To construct the traffic workload model for the industrial-plant scenario, we consider an 802.11 network with a single access point, serving as a backbone for a sensor network deployed for monitoring and actuation purposes. The WLAN terminals are fixed and forward sensor data towards a data-management center, that is, towards the WLAN AP, and control data to sensors and actuators. To define the traffic workload, we follow the use case of [28]. WLAN terminals cover similar areas in the industrial plant. During normal conditions they receive periodic messages from the sensors. These messages are first merged into larger packets at the WLAN terminals and then forwarded to the WLAN AP. Bursty messages arrive from or transmitted to the AP during alarms.

Consequently, we construct the multi-layer traffic workload model as follows. A fixed number of sessions is used for modeling the stationary WLAN terminals. We define two kinds of flows for each session. Flows representing the periodic monitoring traffic to the WLAN AP exhibit uniform flow inter-arrival times with small support, to reflect the random delay in the sensor packet aggregation process. The bursty traffic of alarms is represented by flow inter-arrival times according to heavy-tailed Generalized Pareto distribution. For both of the flow types we consider uniformly distributed flow size, representing the slightly different amount of information forwarded to or received from the data-management center. Similarly, we consider uniform distribution with small support for the packet sizes and packet inter-arrival times within the flows. Table 4 summarizes the traffic workload parameters for the industrial-plant WLAN.

4. Simulations and validation tools

In this Section we describe the simulator framework as well as the analytic tools that we employ for the validation of the semi-Markovian channel occupancy model.

4.1. Simulation framework

The considered networking scenarios are simulated on the *NSmiracle* [29] platform. In all cases an IEEE 802.11b-compliant network is built, employing existing NSmiracle Channel, PHY, MAC and higher layer modules.

Table 4
Industrial-plant WLAN. Traffic workload model parameters.

Traffic object	Distribution Parameters	
Session		
Session number	Fixed	$N \in (5, 20)$
Flow		
Flow inter-arrival time (s)		
– Monitoring	Uniform (α, β)	$\alpha = 5, \beta = 90$
– Alarm occurrence & actuation	gPareto (ξ, σ)	$\xi \in (0.25, 0.5)$, $\sigma \in (5, 10)$
Flow size (bytes)	Uniform (α, β)	$\alpha = 512, \beta = 1536$
Packet Statistics		
Packet size (bytes)	Uniform (α, β)	$\alpha = 512, \beta = 1024$
Packet inter-arrival time (ms)	Uniform (α, β)	$\alpha = 1, \beta = 10$

We consider a single access point area with a radius of 100 m, inside which 802.11 terminals are uniformly distributed. The channel propagation model depends on the considered networking scenario. For the campus WLAN scenario we assume a path-loss-based channel model with a moderate exponent value, $\theta = 2.5$. The model is enhanced by log-normal shadowing with standard deviation $\sigma_{\text{shd}} = 10$ dB. For the in-door scenarios of conference-hall and industrial-plant the channel attenuation is modeled considering a site-general Indoor ITU model [30] with power decay $N = 38$, and zero floor-penetration attenuation. Unless otherwise noted, we use a simplified PHY layer, where the user terminal transmits with a fixed data-rate of 11 Mbps for the industrial-plant scenario, while for the campus and conference-hall WLAN cases the rate is inversely proportional to the distance from the AP, in the 1–11 Mbps interval, to reflect adaptive rate control. In all cases the MAC follows the IEEE 802.11 standard with an IPv4 network layer on the top. Flows are transmitted with TCP or UDP according to their traffic type. The Session module of the simulator implements the multi-layer traffic workload model introduced in Section 3. The directions of the flows are selected randomly, for all cases, apart from the monitoring traffic flows in the industrial-plant scenario, which are always directed from the terminals to the AP.

Furthermore, we implemented a simple protocol stack for a *sensing device*. This device is responsible for continuously measuring the spectrum activity, for collecting a sequence of M samples of idle channel durations, and for building the *empirical* distribution function, $F_{Ie}(t; M)$. It also estimates the parameters of the semi-Markovian channel occupancy model (p, ξ, σ) based on the collected idle period sequences, applying the MLE estimation algorithm in [25]. The outcome of the estimation is the *analytic* distribution function $F_I(t; \xi, \sigma, p)$ of the idle time duration.

4.2. Analytic validation tools

Our goal is to validate the semi-Markovian channel occupancy model described in Section 2.2, using the statistics of the idle channel durations collected from the simulations with the multi-layer traffic workload model.

We assess the *goodness-of-fit* by evaluating the *D-value* of the *Kolmogorof–Smirnoff* test, defined as the maximum absolute difference between the analytic and empirical idle period distribution functions:

$$D = \sup_{\tau \in \mathcal{T}} |F_I(\tau; \xi, \sigma, p) - F_{Ie}(\tau; M)|, \tag{6}$$

where $\mathcal{T}$ is the set of collected idle periods. A low D-value indicates the good fitting performance of the analytic model. We underline that the D-value is a worst-case metric, as it considers the supremum of the point-wise difference between the two functions, instead of the average.

The goodness-of-fit is, additionally, evaluated by performing *two-sample Kolmogorov–Smirnoff* (K–S) tests [31], where the collected sequences of idle periods, $\mathcal{T}$, are tested against *synthesized* idle period sequences, $\widehat{\mathcal{T}}$, that are generated randomly from the estimated analytic distribution, $F_I(t; \xi, \sigma, p)$. The K–S test assesses the validity of the *null hypothesis*, that is, both idle period series can originate from the same distribution. The test is conducted considering n randomly selected samples of sequences $\mathcal{T}, \widehat{\mathcal{T}}$, denoted as $\mathcal{T}_n, \widehat{\mathcal{T}}_n$, respectively. The evaluation is done by deriving the following two-sample K–S test statistic:

$$K_n = \sqrt{\frac{n}{2}} \sup_{\tau \in \mathcal{T}_n, \hat{\tau} \in \widehat{\mathcal{T}}_n} |F_{Ie}(\tau; n) - F_{Is}(\hat{\tau}; n, \xi, \sigma, p)|, \tag{7}$$

where $F_{Is}(\hat{t}; n, \xi, \sigma, p)$ denotes the empirical distribution of the synthesized random sequence. We assess the null hypothesis by calculating the *p-Value* (p_{KS}) of this test, that is the probability of obtaining a test statistic, K_n, at least as extreme as the one we observe. We reject the null hypothesis at a significance level of $\alpha \in (0, 1)$, if $K_n > K_\alpha$, where K_α is the *critical* value [31] defined as $K_\alpha = k : \Pr\{K_n > k\} < \alpha$.

For a deeper understanding of the results of the goodness-of-fit tests we show typical examples of empirical and fitted analytic distributions as well as *quantile–quantile* (Q–Q) plots.

In addition to the goodness-of-fit study, we evaluate the hypothesis that the lengths of the consecutive idle time periods are uncorrelated, an assumption that is required for the semi-Markovian channel occupancy model. We perform a test of independence, by comparing the *lag-k autocorrelation* values of the obtained sample sequence against a sample sequence that approximates well a *white* noise series with low auto-correlation. We generate the white noise reference by sampling the original idle period sequence with large time gap separations. We repeat the autocorrelation test using different sub-sequences of the original sample series. For each test we record the sign of the difference of the lag-k autocorrelation value of the empirical sample sequence and that of the white noise reference. If the samples of the empirical idle-period sequence are correlated, a large portion of these sign outcomes will be positive. If the samples of the tested time series are indeed independent variables, the autocorrelation values of the empirical sample sequence and the white noise reference bare similar statistical behavior and thus, the sign of their difference gives a Bernoulli (1/2) trial. Therefore, to decide on the validity of the null hypothesis we compute the difference between the positive and negative sign outcomes for all tests. We define the p-value of our test as the probability that a Bernoulli (1/2) sequence gives the same or larger difference as the obtained one. That is, the p-value reflects the probability that the observed correlation metric can occur in a sequence of independent samples. We reject the null hypothesis of independence at a significance level α.

5. Numerical evaluation

5.1. The impact of in-flow characteristics on fitting accuracy

Multi-layer workload models in general and also [7] characterize the higher-level traffic workload in WLANs, while the modeling of in-flow traffic is not considered. Therefore, in the first part of this Section we evaluate the influence of the in-flow traffic characteristics on the fitting performance of the proposed mixed uniform-Pareto idle period distribution. This evaluation will help us conclude whether the selection of a particular application mix with application-specific packet size and packet inter-arrival time distribution can have a significant effect on the fitting performance.

We consider a single high-level workload *configuration*, that is, we generate sessions (session arrival times) and flows (number of flows per session, flow arrival times and flow sizes) considering the campus WLAN model (see Table 1). We use the same workload configuration for all the experiments. For each experiment we select a packet size and inter-arrival time distribution pair according to Table 5, and perform 50 simulation runs (for all but the deterministic packet size, deterministic inter-arrival time case) collecting $M = 10^4$ idle period samples within each run. Table 6 gives the average D-value for each experiment. The D-values are very low for all cases with both random packet sizes and random inter-arrival times, and do not seem to depend on the actual parameter value of the distributions. The mixed Pareto distribution, however, does not fit well the empirical data under short, deterministic inter-arrival times, especially when combined with deterministic packet size. We evaluate the reason of the large D-value with the help of Fig. 3, showing the empirical and analytic idle-period distribution functions for two in-flow configurations.

Fig. 3(a) shows results with Exponential ($\lambda^{-1} = 512B$) packet size and Uniform ($\mu = 10^{-2}, \sigma^2 = 10^{-5}$ s) packet inter-arrival time, exhibiting a D-value = 0.005, while Fig. 3(b) with Deterministic (1024 B) packet size and Deterministic (10^{-2} s) packet inter-arrival time, with D-value = 0.055. Under deterministic packet size and packet inter-arrival time the empirical function, $F_{Ie}(t;M)$ is dominated by either 802.11 back-off periods, or idle periods with a duration around 10^{-2} s, generated between successive packets inside flows. Apparently, the random mixing of the deterministic packet streams does not lead to generalized Pareto distributed idle times in this case.

Based on the results in Table 6 we conclude that the in-flow traffic characteristics may have an impact on the fitting accuracy, and therefore, the validation of the channel occupancy model under real WLAN scenarios must take into account realistic, application-dependent in-flow models.

Table 5
In-flow stochastic models.

Distribution	Parameters		
Packet sizes (bytes)			
Uniform (α, β)	$(64, 192)$, $(256, 768)$, $(768, 1280)$		
Exponential (λ)	128	512	1024
Deterministic	128	512	1024
Inter-arrival time (s)			
Uniform (α, β)	$(0.005, 0.015)$	$(0.05, 0.15)$	$(0.5, 1.5)$
Exponential $(1/\lambda)$	10^{-2}	10^{-1}	1
Deterministic (μ)	10^{-2}	10^{-1}	1

5.2. The impact of the networking scenario on the fitting accuracy

Let us now evaluate the accuracy of the idle time distribution model for the three scenarios introduced in Section 3, that is, the campus, the conference-hall and the industrial-pant WLAN, with the multi-layer traffic workload model parameters summarized in Tables 1, 3, and 4, and the in-flow parameters in Table 2. We simulate the different scenarios with randomized traffic workload parameters and estimate the parameters of the idle time distribution based on an input sequence of $M = 10^4$ idle period samples, allowing for a 10^3-samples warm-up period. Depending on the length of the active and idle time periods this corresponds to different simulation time durations in the order of 10^2–10^3 s. We perform 10^3 simulation runs for each of the scenarios.

Fig. 4 shows the parameter space of the idle time distribution of the semi-Markovian model obtained from the simulations, considering p, the percentage of back-off periods, and the average estimated white space duration $\bar{I}^{(WS)}$ defined in (5). The figures show that the three scenarios present significantly different parameter spaces. Comparing the three scenarios we can identify the significant factors that affect the accuracy of the channel occupancy model and determine the model limitations.

5.2.1. Campus WLAN

Fig. 5(a) shows the empirical cumulative distribution function of the D-value over 10^3 simulation runs. The distribution exhibits a low mean, $\bar{D} = 0.0199$, and moderate variance with 95% of the cases being lower than 0.04, revealing an excellent fitting quality. Fig. 5(b) depicts the quantile–quantile plot for the collected and the synthesized idle series, averaged over all simulation runs. We observe that the quantile–quantile curve follows closely the $x = y$ axis, which indicates that the series of the collected and synthesized idle period samples do come from the same distribution. This is verified as well by the low failing rate of the conducted Kolmogorov–Smirnoff test, shown in Table 7. The average p-value of the test is $p_{KS} = 0.5714$, with a coefficient of variation of $C_{p_{KS}} = 0.0923$, while its failing rate is 7.01%, for the standard significance level, $\alpha = 5\%$.

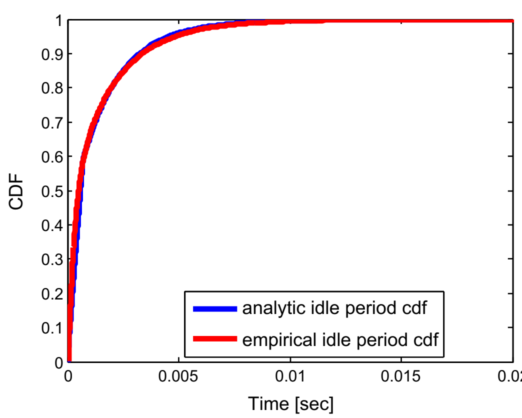
Table 6
Expected *D*-value for the various in-flow configurations.

Packet size (bytes)		Packet inter-arrival times (s)								
		Uniform			Exponential			Deterministic		
		10^{-2}	10^{-1}	1	10^{-2}	10^{-1}	1	10^{-2}	10^{-1}	1
U	128	.0122	.0151	.0171	.0205	.0165	.0169	**.0945**	.0220	.0344
	512	.0093	.0115	.0177	.0180	.0136	.0176	**.0943**	.0508	.0291
	1024	.0248	.0144	.0181	.0101	.0144	.0171	**.0899**	.0196	.0356
E	128	.0169	.0218	.0166	.0170	.0229	.0163	**.0706**	.0287	.0303
	512	.0243	.0217	.0168	.0294	.0221	.0167	**.0574**	.0418	.0280
	1024	.0251	.0409	.0163	.0454	.0416	.0157	**.0912**	.0476	.0314
D	128	.0140	.0097	.0183	.0114	.0094	.0179	**.0676**	**.0665**	.0461
	512	.0110	.0105	.0185	.0085	.0092	.0181	**.1486**	**.0817**	.0455
	1024	.0199	.0113	.0178	.0079	.0111	.0178	**.1733**	**.0362**	.0448

Even though the average fitting quality is satisfactory, we would like to identify the scenarios when the mixed uniform-Pareto distribution fails to adequately model the idle periods. Therefore, we first classify the simulation results according to the experienced WLAN *channel load*, defined as the percentage of the time the channel is in active state. Fig. 6(a) demonstrates that the channel load affects the *D*-value. The fitting accuracy is low in light-loaded cases, and the fitting is rather weak even at high load values. This indicates that the idle period distribution model is more suitable for capturing the channel occupancy statistics for moderate load cases. The empirical density function of WLAN channel load is shown in Fig. 6(b). We observe that the cases of extreme channel load, which result in poor model fitting quality, are relatively rare, and therefore we can conclude that the proposed analytic model is, in general, sufficient for channel occupancy modeling in campus WLANs.

Fig. 7 shows the fitting quality of the semi-Markovian model, when the WLAN terminals have higher transmission rate capabilities, considering the same traffic as for Fig. 6. We consider fixed, distance-dependent transmission rate in the 1–54 Mbps range in Fig. 7(a) and (b) and per packet dynamic rate adaptation in the same range in Fig. 7(c) and (d). In both cases the fitting quality follows a similar *D*-value – WLAN load trend, as in Fig. 6, indicating that the transmission rates of the terminals do not significantly affect the characterization of the idle channel period durations. The fitting quality improves slightly under per packet rate adaptation due to the additional randomization of active and, consequently, of the idle period lengths. Due to the high transmission rates we observe a few simulation runs where the WLAN channel load is relatively low (1–10%); in these cases the fitting quality is again relatively weak.

In Fig. 8 we depict the relation between the *D*-value and the number of *active* sessions, that is, all the sessions that arrive and transmit during the simulation run. Fig. 8(a) suggests that the number of sessions affects the fitting quality of the model, with low number of sessions leading to low fitting quality. As low number of sessions usually means low load, we evaluate whether the load or the number of sessions has dominant effect. In Fig. 8(b) we plot the *D*-value with respect to the number of sessions, restricting the study for the cases of low or heavy channel load. As simulation runs with very low and very heavy load are rare, we select load regions $\leq 25\%$ and $\geqslant 45\%$ to include a reasonable number of runs. We observe that under



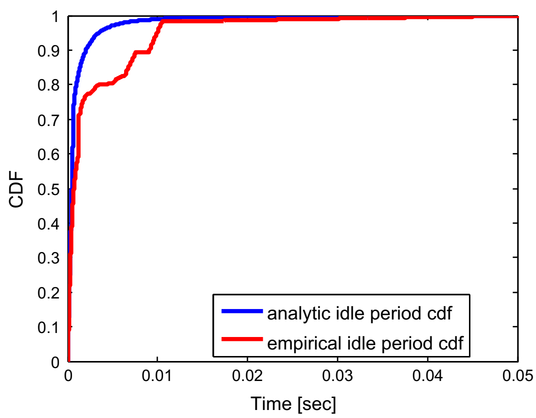


Fig. 3. Examples of fitting performance for different in-flow characteristics: (a): *D*-value = 0.005, Exponential ($\lambda^{-1} = 512$ B) packets-size, Uniform ($\mu = 10^{-2}, \sigma^2 = 10^{-5}$ s) packet inter-arrival time. (b): *D*-value = 0.055, Deterministic (1024 B) packet-size, Deterministic (10^{-2} s) packet inter-arrival time.

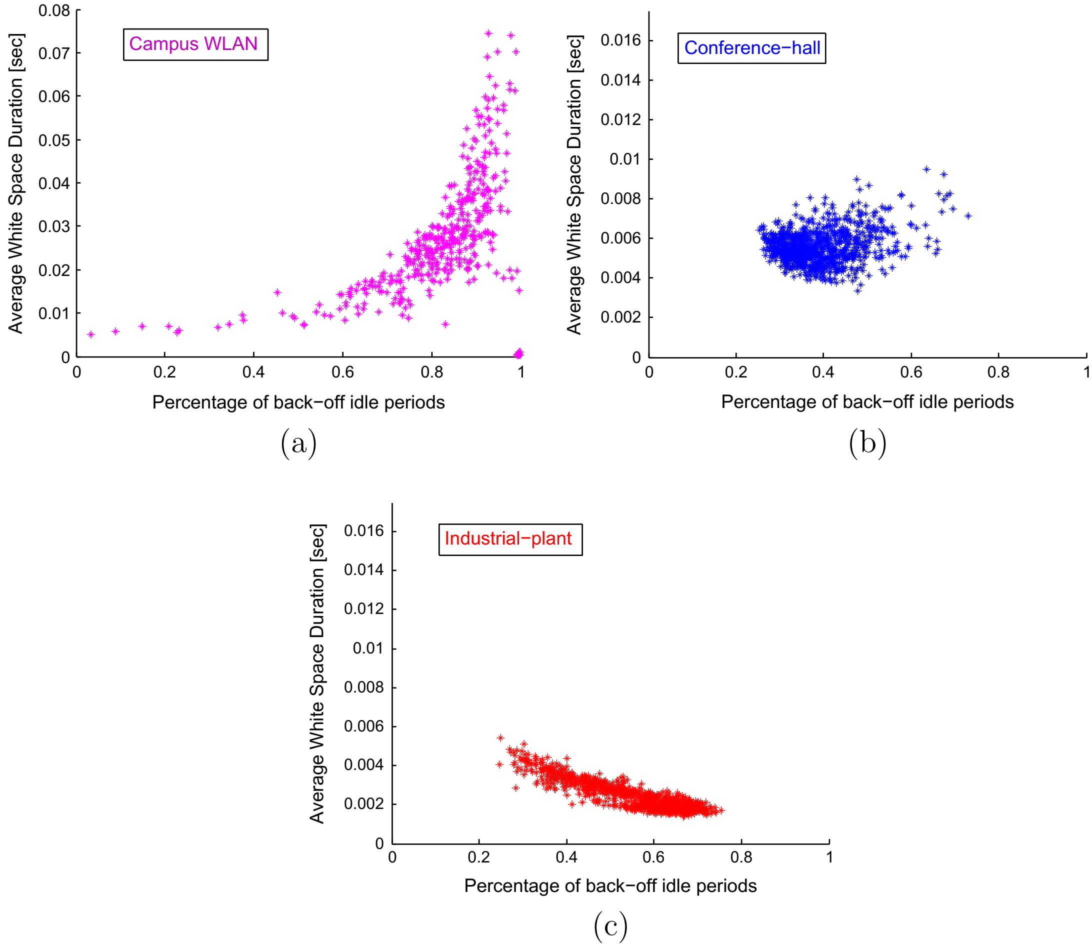


Fig. 4. The resulting parameter space for the idle spectrum period distribution for the three networking scenarios.

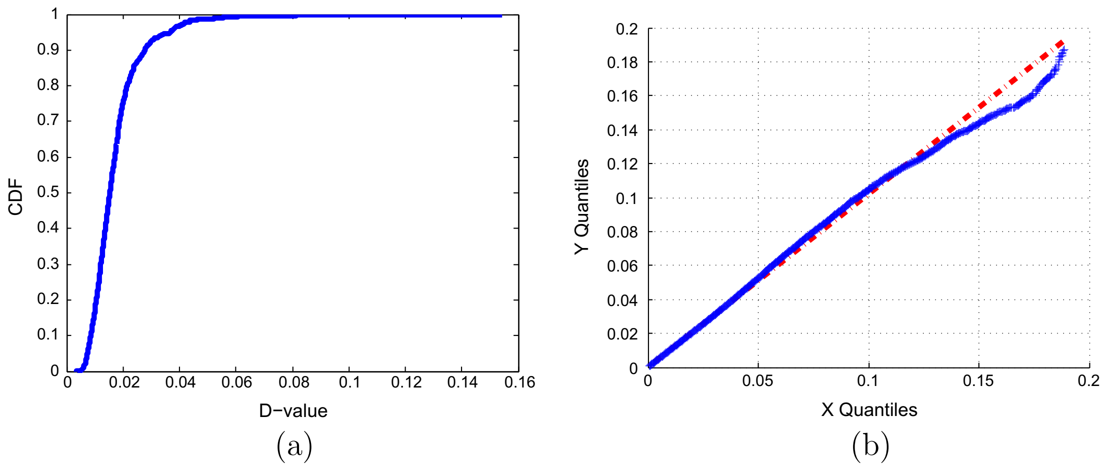


Fig. 5. Campus WLAN. (a) Empirical distribution function of the *D*-value. (b) Q–Q plot for the empirical and synthesized idle period series over all simulation runs.

similar channel load, the number of the active WLAN sessions has little impact on the fitting quality, unless both the network load and the number of sessions are very low. We evaluate the reason of the high *D*-value in this region by comparing the empirical and analytic idle period CDFs for a simulation run in Fig. 9(a). As we see, the

Table 7
Summary of goodness-of-fit evaluation for the considered networking scenarios.

	D-value	p_{KS}	$C_{p_{KS}}$	$P\{p_{KS} \leqslant 5\%\}$
Campus	0.0199	0.5714	0.0923	0.0701
Conference-hall	0.0088	0.5335	0.0214	0.0467
Industrial-plant	0.0334	0.4029	0.2103	0.0934

empirical CDF first increases rapidly, but then shows heavy-tail characteristics. Due to the low number of active sessions, each with low load, flows rarely overlap in time, and the idle period distribution is determined by the in-flow characteristics, that cannot be captured by the generalized Pareto distribution. Fig. 9(a) shows, additionally, that in low-load cases the percentage of entirely idle beacon intervals may be significant, therefore, modeling the white-spaces with the truncated density in (4) is essential for achieving high accuracy.

Fig. 10 shows the relation between the D-value and the level of traffic dispersion among the active sessions. We define the *normalized traffic dispersion*, η, as:

$$\eta = \sum_{i=1}^{N} \left| l_i - \frac{1}{N} \right|,$$

where N is the number of active sessions in the measurement window, and l_i indicates the ratio of total active duration due to session $i, \sum_{i=1}^{N} l_i = 1$. Thus, η is zero under completely balanced traffic, low when the network traffic is roughly evenly distributed among the sessions, while high η values indicate unbalanced traffic load. Fig. 10(a) suggests that unless it is very high, the traffic dispersion does not have an impact on the fitting quality. However, as it is shown in Fig. 10(b), dispersion has different effects in low and in high load regions. For low or moderate channel load the fitting accuracy degrades with increasing traffic dispersion. As at high dispersion the large part of the load is generated by a subset of the sessions, the reason of the low fitting quality is the same as for the low load – low session number case, that is, the idle time distribution is determined directly by the in-flow characteristics. Under high channel load the fitting quality is generally good, apart from the case of very balanced load. We investigate the reason of the high D-value in this case in Fig. 9(b). According to the figure, the empirical and analytic CDFs do not fit at the back-off period interval. High channel load together with balanced traffic means that many sessions access the wireless channel concurrently. This leads to high level of contention and thus to the exponential increase of the user back-off window. The assumption of uniformly distributed back-off period length cannot hold in this case, which degrades the performance of the fitting.

Finally, we investigate the impact of transmission errors on the fitting quality. Fig. 10(c) depicts the relation between the D-value and the MAC *error rate*, that is the percentage of packet transmissions, for which an error has occurred, including a CCA failure, that is, terminated transmission attempt after maximum back-off, collision due to hidden terminals, or error in decoding data or ACK packet due to bad channel conditions. Clearly, the fitting quality degrades at high error rate. To determine the reasons behind the fitting performance degradation, we plot in Fig. 10(d) the relation between the MAC error rate and the fitting quality for different WLAN load regions. Under low or moderate channel load, the increased MAC error rate, mainly, due to bad channel conditions, does not have an effect on the fitting quality. This is because the packet retransmission scheme of 802.11 does not affect the distribution of the back-off periods and it does not significantly shrink the channel white spaces. As the channel occupancy becomes high, the MAC error rate increases mainly because of CCA failures and collisions due to hidden-terminals, and the increasing D-value reflects that the fitting quality is low due to the non-uniform distribution of the back-off periods, as a consequence of the high-load itself.

We can conclude, that in the scenario of a campus WLAN, the mixture distribution, proposed to characterize the idle period lengths, is accurate for the typical cases with moderate channel load. We have observed that the fitting quality is worse at low load and at low number of

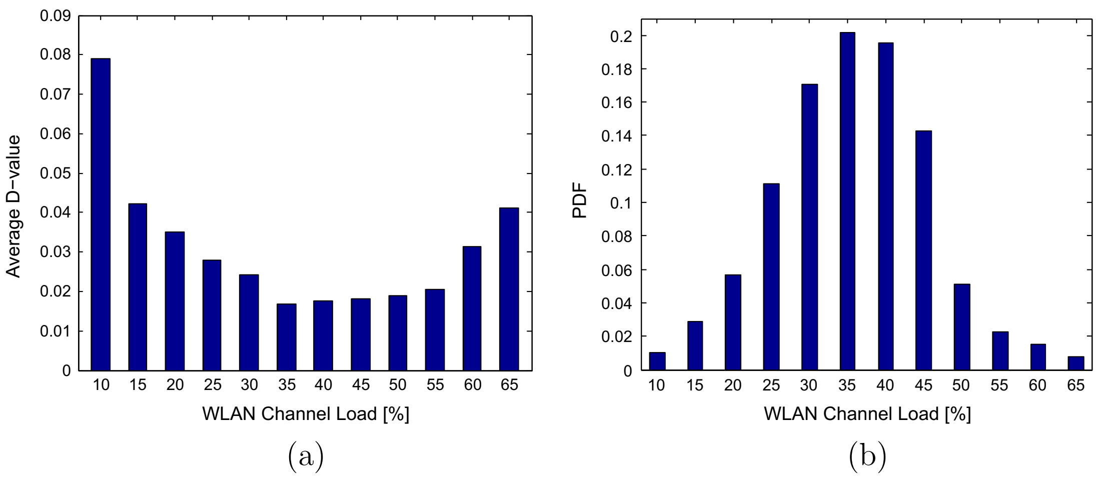


Fig. 6. Campus WLAN. (a) Average D-value with respect to WLAN channel load and (b) empirical density function of the channel load.

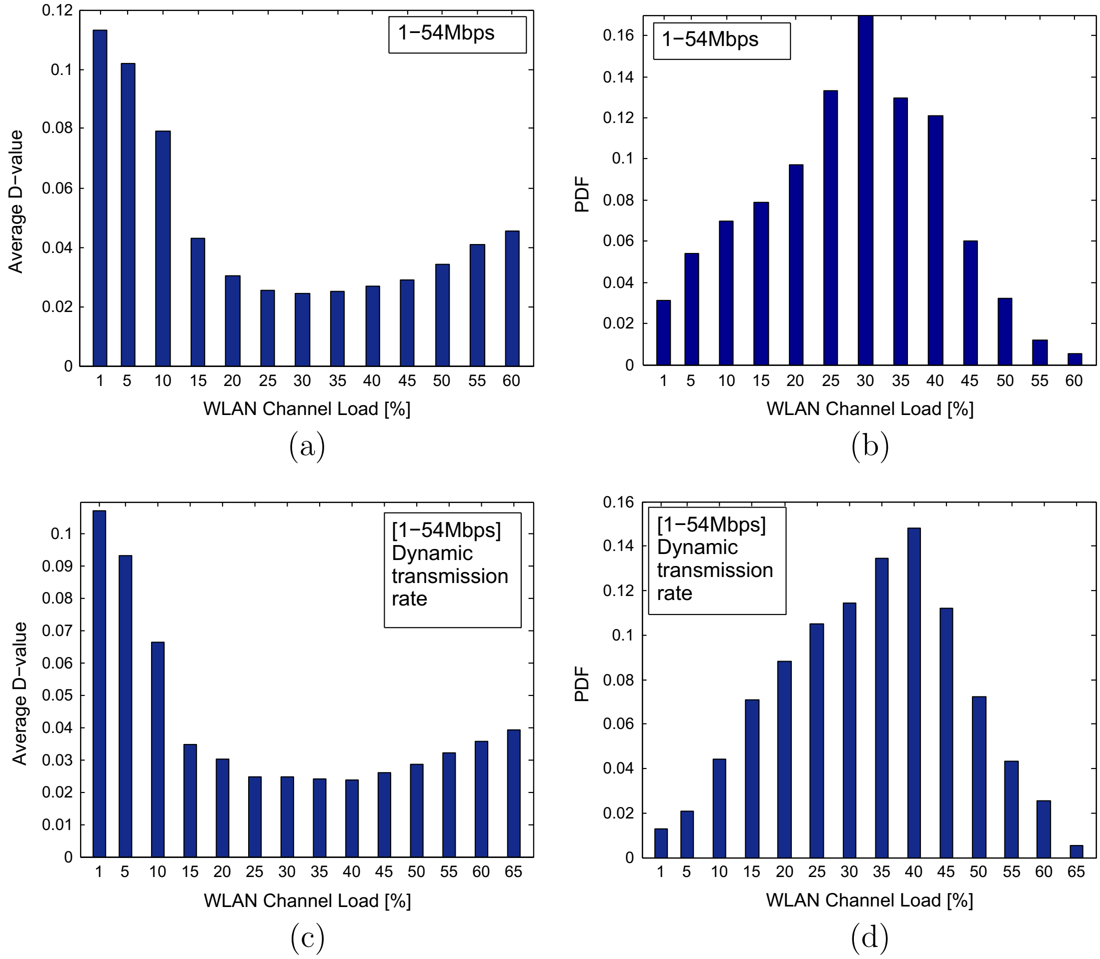


Fig. 7. Campus WLAN. (a) and (c) Average *D*-value with respect to WLAN channel load, (b) and (d) empirical density function of the channel load for the case of 802.11 terminals with (a) and (b) higher transmission rates and (c) and (d) the case of 802.11 terminals with dynamic rate adaptation.

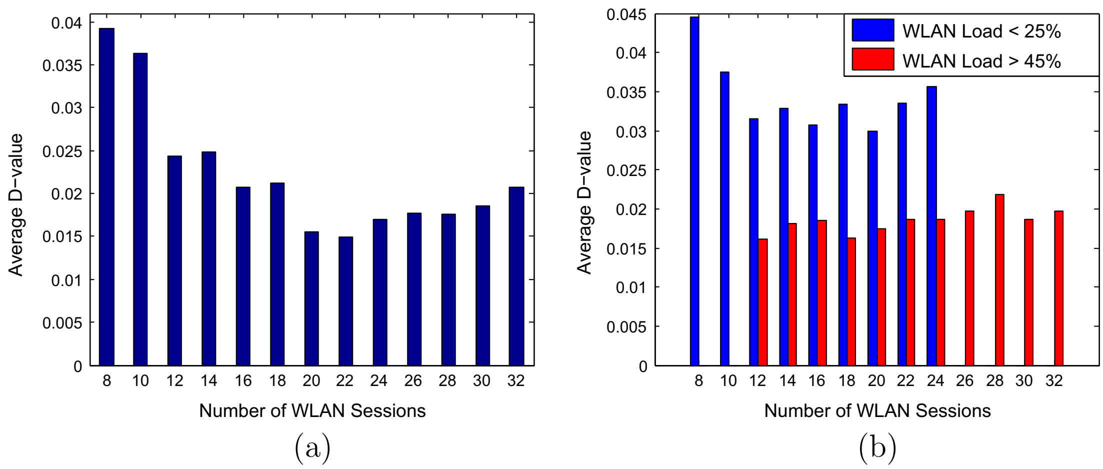


Fig. 8. Campus WLAN. (a) Average *D*-value with respect to the number of sessions considering (a) all cases and (b) cases with load below 25%, and above 45%.

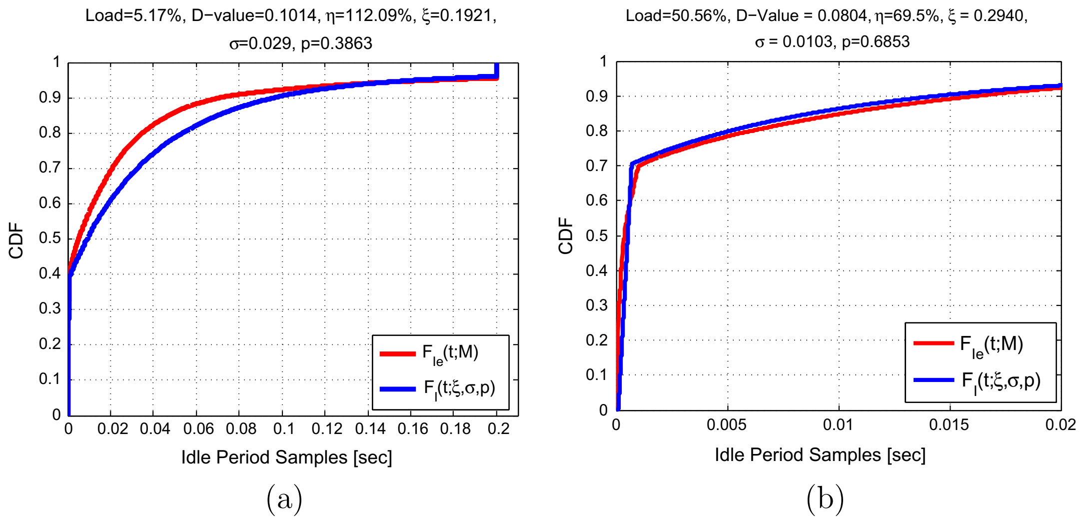


Fig. 9. Campus WLAN. Comparison between the empirical and the analytic CDF of the idle channel period durations for a single simulation run, considering (a) low WLAN load (5.17%), and (b) high WLAN load (50.56%) with low dispersion ($\eta = 69.5\%$).

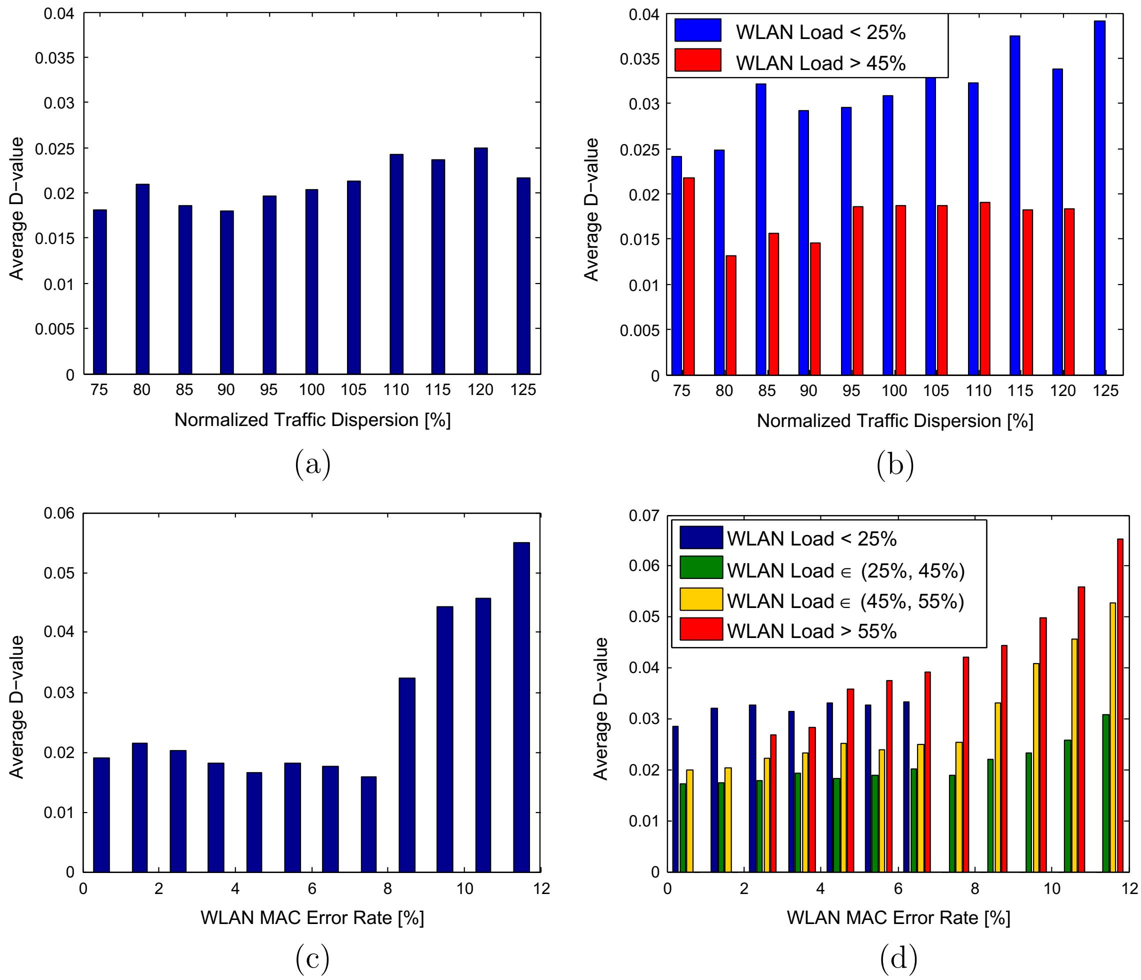


Fig. 10. Campus WLAN. Average *D*-value with respect to the normalized traffic dispersion among the WLAN users considering (a) all cases and (b) cases with load below 25%, and above 45% and with respect to the MAC error rate, considering (c) all cases and (d) cases with different WLAN load regions.

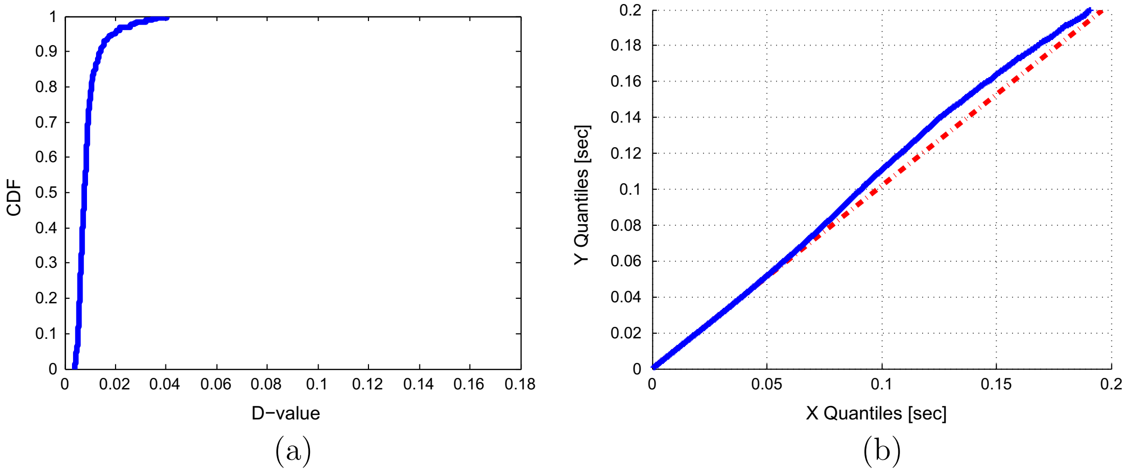


Fig. 11. Conference-hall. (a) Empirical distribution function of the *D*-value. (b) Q–Q plot for the empirical and synthesized idle period series over all simulation runs.

sessions or at high level of traffic dispersion among the sessions, because in this case very few packet streams are aggregated and the in-session flow characteristics determines the idle period distribution. The fitting quality can be low even at very high, well balanced load, and under high MAC error rate, as in these cases the back-off period model is not accurate enough.

5.2.2. Conference-hall WLAN

As Fig. 4(b) shows, the conference-hall WLAN exhibits a set of differences compared to the campus WLAN scenario, with significantly lower variation of the average idle period length and of the percentage of back-offs across the simulation runs. From Fig. 11(a) we can see, that *D*-values, in general, are lower than in the campus WLAN case. As given in Table 7, their average is $\overline{D} = 0.0088$ and this is reflected in the respective Kolmogorov–Smirnoff test where the average *p*-value is 0.5335, but with a low coefficient of variation of 0.0214, and, therefore, a low null hypothesis rejection rate equal to 4.67%. Fig. 11(b) with the Q–Q plot of the collected and synthesized idle period series shows the same good accuracy, with a good fit up to very high time values.

As for the campus scenario, let us evaluate how the load, the number of sessions, the traffic dispersion and the MAC error rate affect the fitting accuracy. Fig. 12(a) depicts the relation between the network load and the resulting *D*-value. Similarly to the campus WLAN scenario, the fitting accuracy is weaker at very low network load, but after that the *D*-value becomes low and independent of the load. Moreover, as shown in Fig. 12(b), the WLAN load ranges, mostly, between 15% and 35%, with the majority of the simulation runs showing an average load of 20–25%, a range that is significantly shorter than that of the campus WLAN case. The fitting quality is very good under the typical load levels, showing that the analytic idle time model is adequate in the conference-hall scenario. As Fig. 13 shows, the fitting accuracy in the conference-hall

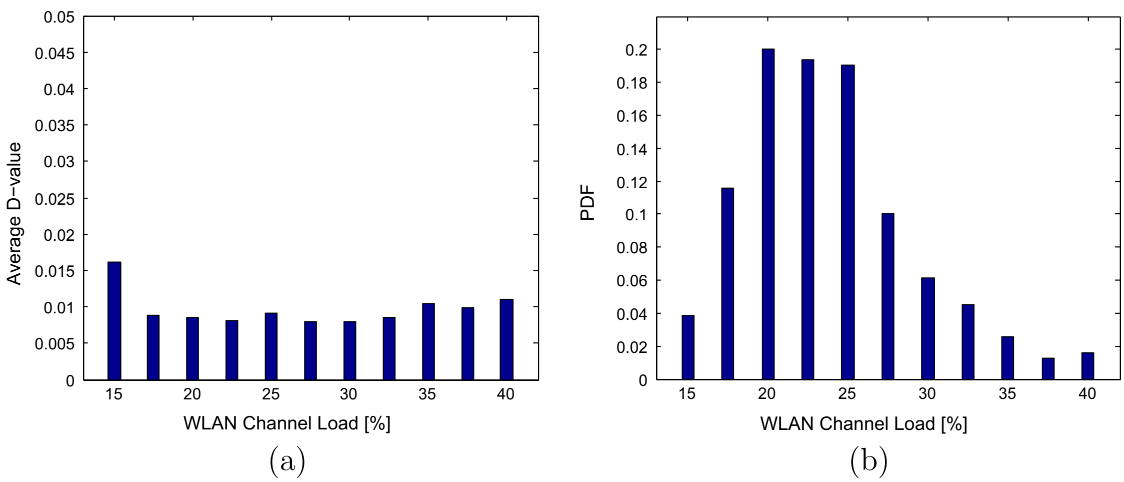


Fig. 12. Conference-hall. (a) Average *D*-value with respect to WLAN channel load, and (b) empirical density function of the channel load.

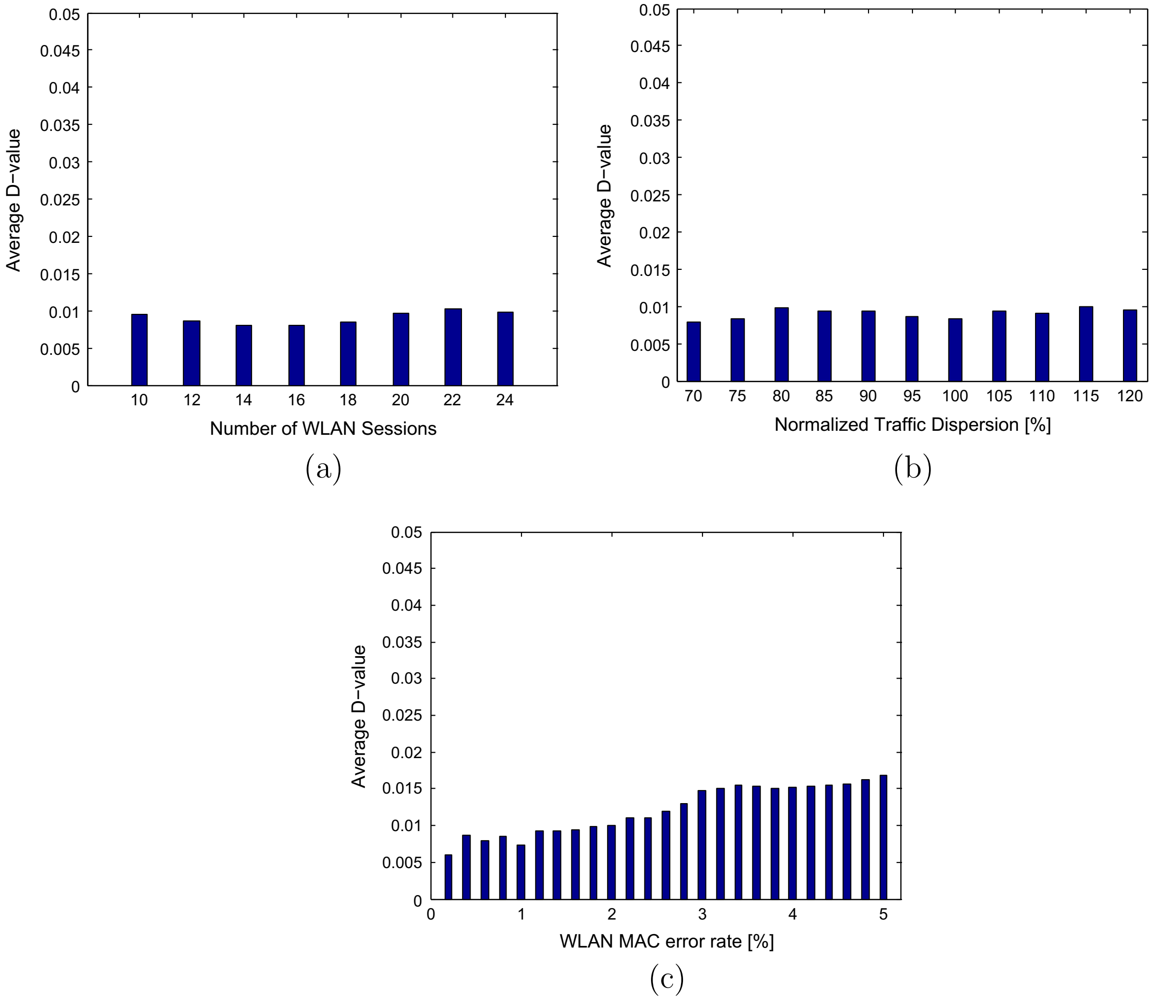


Fig. 13. Conference-hall. Average *D*-value with respect to (a) the number of sessions, (b) the normalized traffic dispersion among the WLAN users, and (b) the MAC error rate.

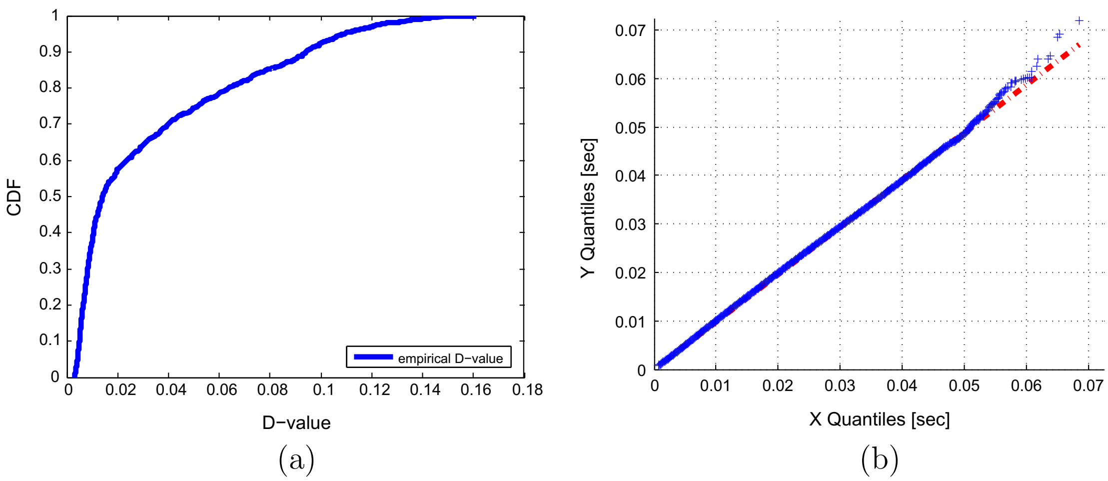


Fig. 14. Industrial-plant WLAN. (a) Empirical distribution function of the *D*-value. (b) Q–Q plot for the empirical and synthesized idle period series over all simulation runs.

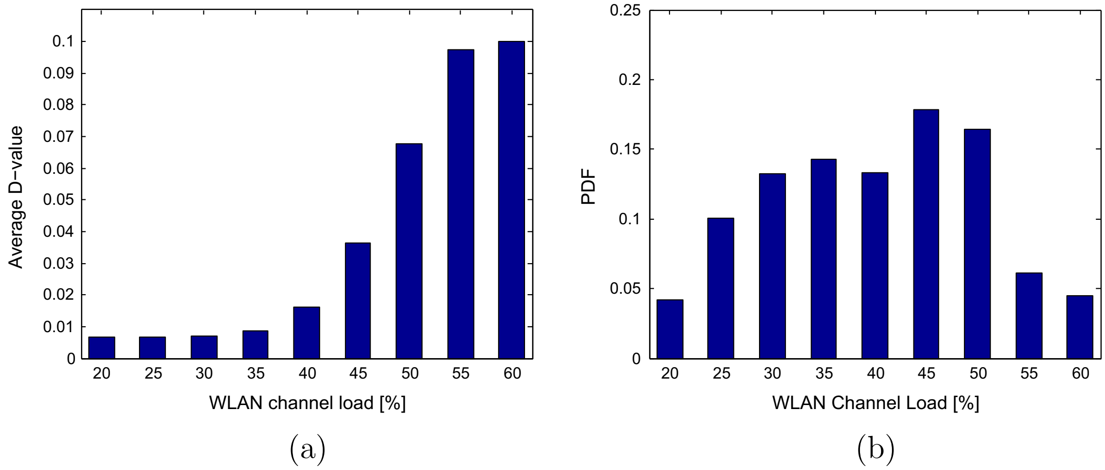


Fig. 15. Industrial-plant WLAN. (a) Average *D*-value with respect to WLAN channel load, and (b) empirical density function of the channel load.

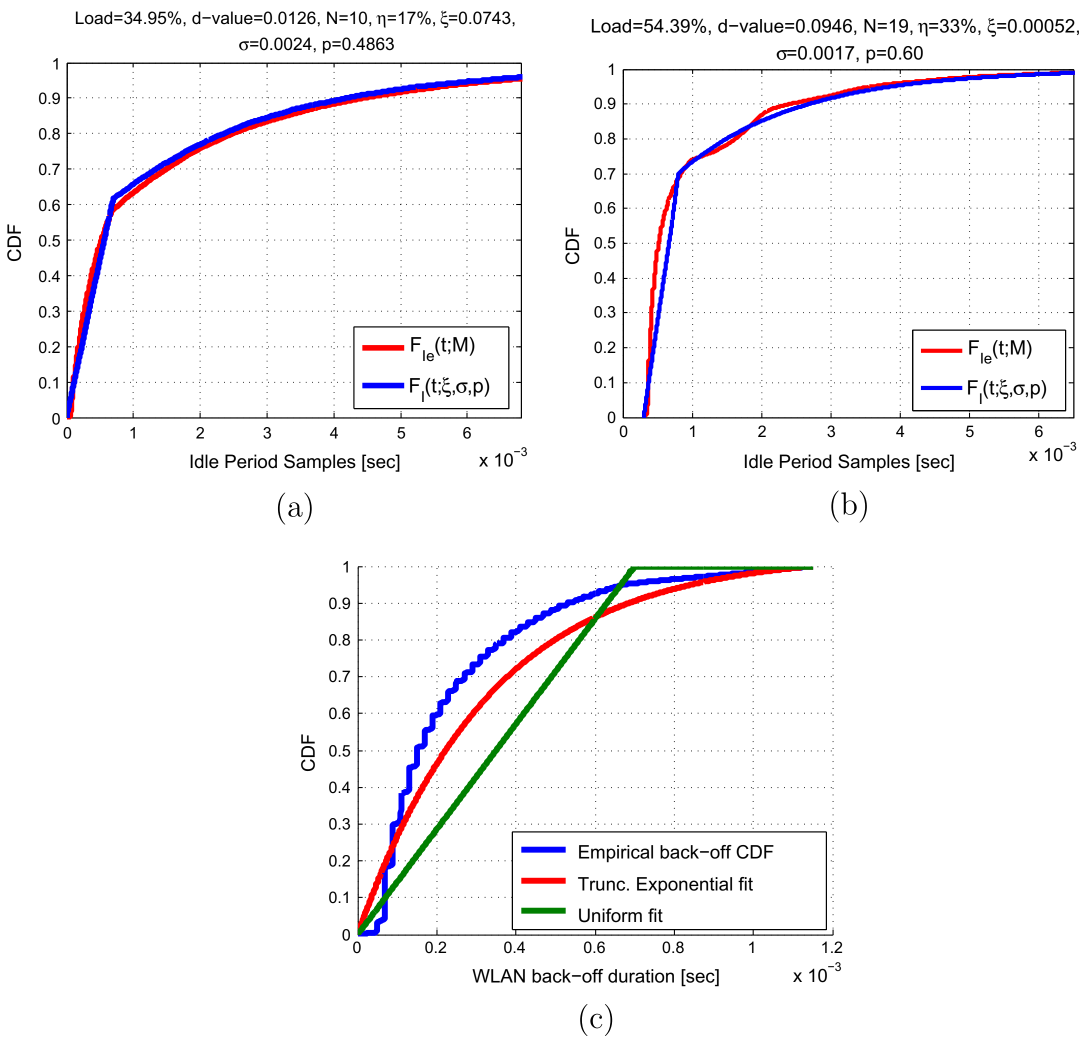


Fig. 16. Industrial-plant WLAN. Comparison between the empirical and the analytic CDF of the idle channel period durations for a single simulation run, considering (a) moderate WLAN load (34%) and (b) high WLAN load (54.39%). (c) Compares the exponential-based fitting of the back-off WLAN periods, as opposed to the employed uniform distribution.

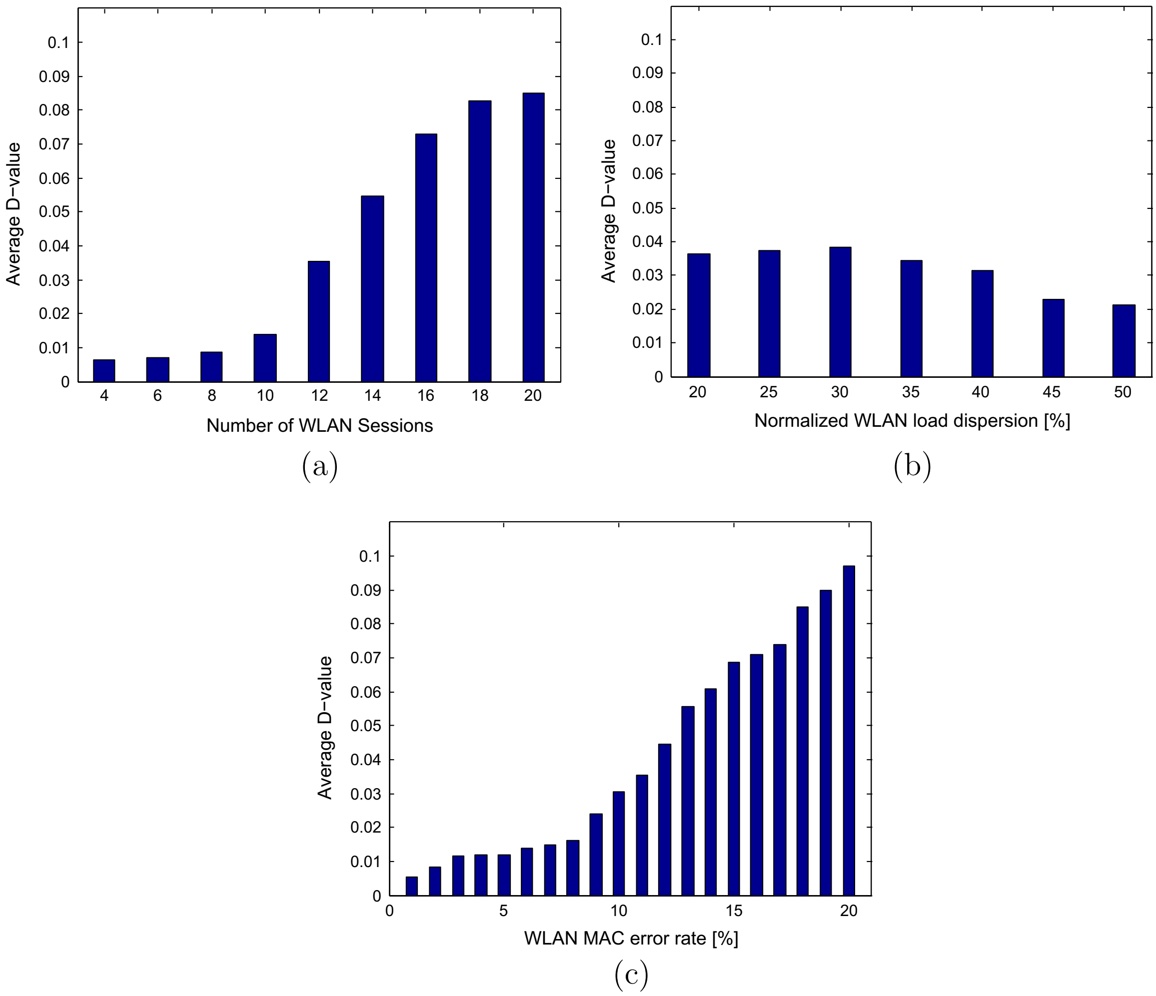


Fig. 17. Industrial-plant WLAN. Average *D*-value with respect to (a) the number of sessions, (b) the normalized traffic dispersion among the WLAN users, and (c) the MAC error rate.

case is not affected by the number of sessions or by the load dispersion. Higher MAC error rates lead to a slightly decreased fitting quality. However, contrary to the campus WLAN scenario, both the D-value and the error rates are relatively low. Error rates are low because packet errors and hidden terminals are not as frequent due to better channel conditions – as described in Section 4.1 – while CCA failures are not present as the load does not reach very high values. Consequently, the low error rates do not significantly affect the fitting accuracy, as discussed earlier in Fig. 10(d).

Comparing the conference-hall results to the campus WLAN ones, we can see that the *D*-values are lower under similar channel load values, and are less affected by the level of traffic dispersion. This good behavior follows from the in-flow characteristics in the conference-hall scenario. As we can see in Table 3, the traffic is rather homogeneous, dominated by Web flows, each of them with relatively long, exponentially distributed packet inter-arrival times. The mixed uniform-Pareto distribution seems to fit very well the idle-time distribution resulting from the aggregation of a high number of such flows.

We can conclude that the proposed channel occupancy model is accurate for the conference-hall WLAN scenario.

5.2.3. Industrial-plant WLAN

As illustrated in Fig. 4(c) the channel occupancy in the case of the industrial-plant WLAN resembles the conference-hall case, with short idle period durations and low variations in the percentage of back-off idle periods. As shown in Table 7, the average *D*-value is 0.0321, higher, compared to the campus and conference-hall cases. The average *p*-value of the K–S goodness-of-fit test is, consequently, the lowest among all cases (0.4029), which, together with a relatively high coefficient of variation (0.2103) results in the highest null hypothesis rejection rate, 9.34%. However, the CDF of the D-value, shown in Fig. 14(a) and the Q–Q plot in Fig. 14(b) still show a good fitting accuracy.

Repeating the evaluation process, we investigate whether the channel load, the number of sessions, the traffic dispersion, or the MAC error rate affect the fitting quality. As shown in Fig. 15(a), contrary to the conference-hall case, the fitting accuracy decreases significantly in cases of

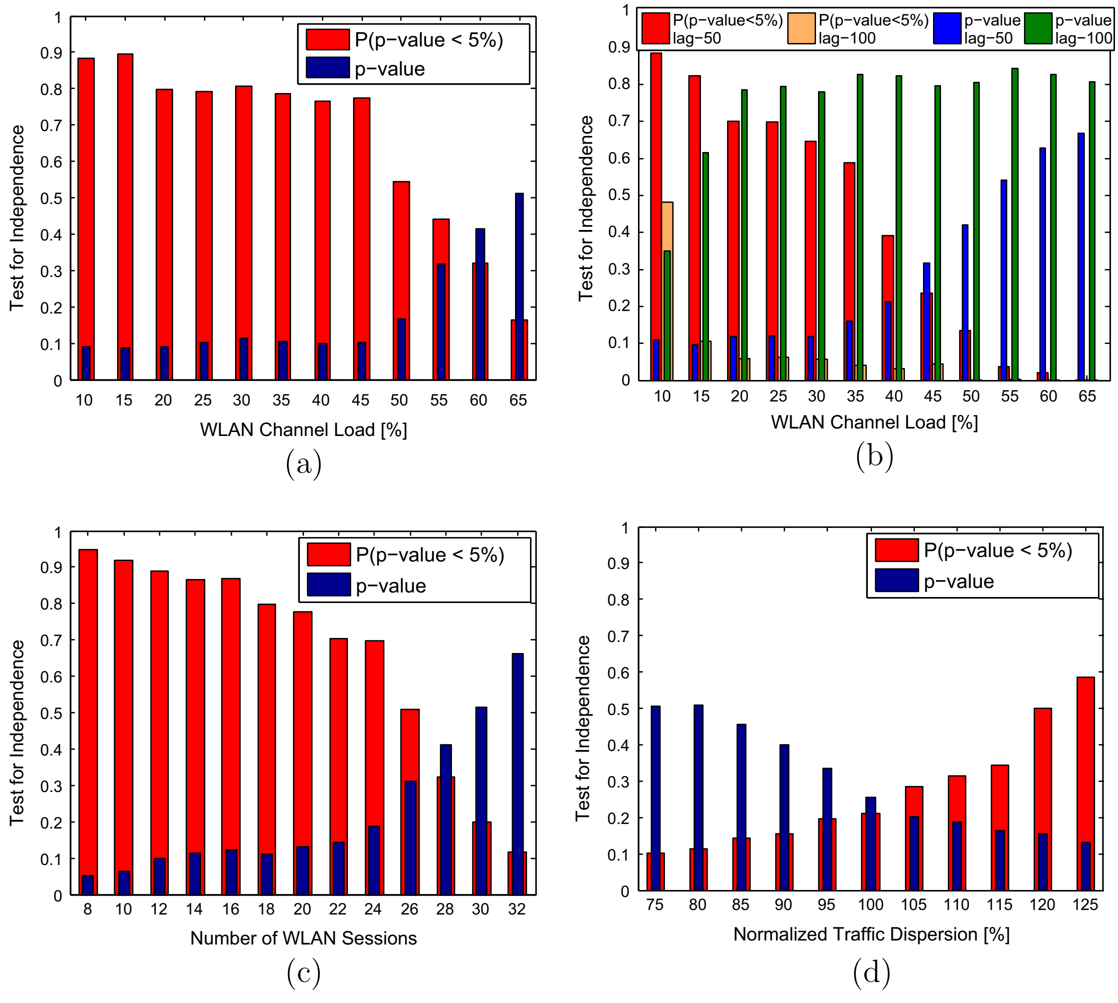


Fig. 18. Campus WLAN. Test for lag-1 independence for the idle period process. Average *p*-value and the percentage of *p*-values below the 5% significance level, with respect to (a) the WLAN channel load, (c) the number of WLAN sessions, and (d) the normalized traffic dispersion among the sessions. (b) Lag-50 and lag-100 independence tests.

high channel load (Fig. 15(a)), moreover, these cases are rather frequent, as shown in Fig. 15(b). We compare the idle period distribution for moderate and high load cases in Fig. 16. Moderate channel load leads to good fitting quality (Fig. 16(a)), even though, due to the periodic monitoring traffic, the idle distribution, $fI_e(t;M)$ is upper bounded, that is, clearly, non heavy-tailed. Fig. 16(b) evaluates the reason for the high D-value under high load. The empirical CDF diverges from the analytic one at very low idle period values, showing that, due to the high contention level, the back-off periods are again non-uniform, as for the Campus case on Fig. 9(b). We show the empirical distribution of the idle back-off period sequence in Fig. 16(c). Under high contention, back-off periods extend above the maximum value, $\alpha_{bk} = 0.7$ ms, considered in our model estimation process, which results in fitting error above α_{bk}. However, the larger gap is at the lower time values, where the back-off period distribution is clearly non-uniform. To keep the estimation process feasible, we propose to fit the back-off period distribution with an exponential-like function, which can resemble the linear combination of increasing back-off period length, but requires the estimation of a single parameter. Fig. 16(c) compares the fitting performance of a left–right truncated exponential distribution with the standard uniform density, indicating, clearly, that the former is more capable of capturing the real behavior of the 802.11 back-off periods. We have to notice, however, that the estimator algorithm must be performed after the estimation of the mixture parameter *p*, which decreases the achievable accuracy. Still, the empirical CDF diverges from the analytic one even for low white-space period values. We believe that consecutive contention periods under the rather bursty traffic of the industrial plant delay the transmission of consecutive packets for several time-slots, causing the short white-spaces to disappear.

As shown in Fig. 17(a), the fitting accuracy degrades as well with the number of WLAN sessions. However, as in this case the number of sessions and the network load are strongly correlated, the reason for the high D-value is

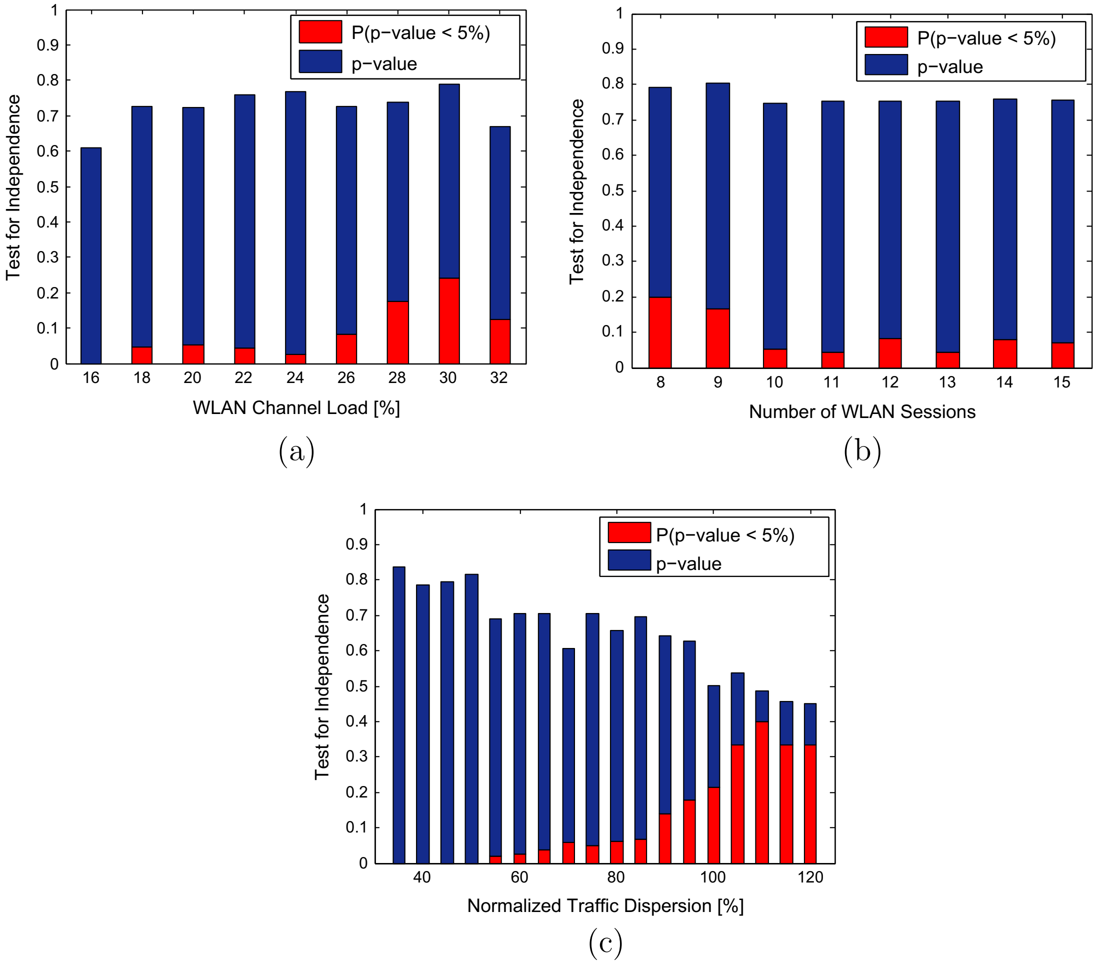


Fig. 19. Conference-hall. Test for independence for the idle period process. Average *p*-value and the percentage of *p*-values below the 5% significance level, with respect to (a) the WLAN channel load, (b) the number of WLAN sessions, and (c) the normalized traffic dispersion among the sessions.

the same as at the high-load case above. The normalized traffic dispersion is low ($\eta \in (20$–$50\%)$), as the monitoring application generates similar traffic on all WLAN sessions (Fig. 17(b)), and has little effect on the model accuracy. According to Fig. 17(c), the MAC error rates can be high, compared to the campus and conference-hall cases, due to the higher transmission rate (11 Mbps), that increases the bit error rate probability at larger terminal-AP distances. The fitting accuracy degrades with increasing error rate, as we have seen in the previous scenarios.

To conclude, in the industrial-plant scenario the channel occupancy model is accurate in many cases, but the fitting quality depends heavily on the channel load and the MAC error rate. The mixed uniform-Pareto idle period distribution can be applied with high confidence for networks with up to moderate load and less than 10% MAC error probability.

5.3. Evaluation of the Markovian assumption

Finally, we investigate whether the durations of the successive WLAN channel idle periods are independently distributed random variables, which is a fundamental assumption for the semi-Markovian occupancy model. For the considered three scenarios we evaluate the effect of the channel load, the number of sessions and the traffic dispersion, by applying the test for independence, described in Section 4.2. We consider lag-1 autocorrelation, unless otherwise noted.

For each scenario we conduct 500 simulation runs with randomized traffic workload parameters, thus resulting in different channel load, number of sessions and traffic dispersion values. For each simulation run we collect a sequence of $4 \cdot 10^4$ idle period samples. The long sequence is divided to 100 intervals of 400-samples length. We construct the white noise reference sequence by selecting one sample from each interval. In each interval we select randomly one of the first 100 samples, so as to guarantee both randomness and a minimum time separation of 300 samples. Then we perform the lag-1 test of independence within each of the 400-sample intervals, considering the sub-sequence of the first 100 idle period samples. We finally calculate the *p*-value of the test over the 100 intervals. Recall, the *p*-value reflects the probability that the

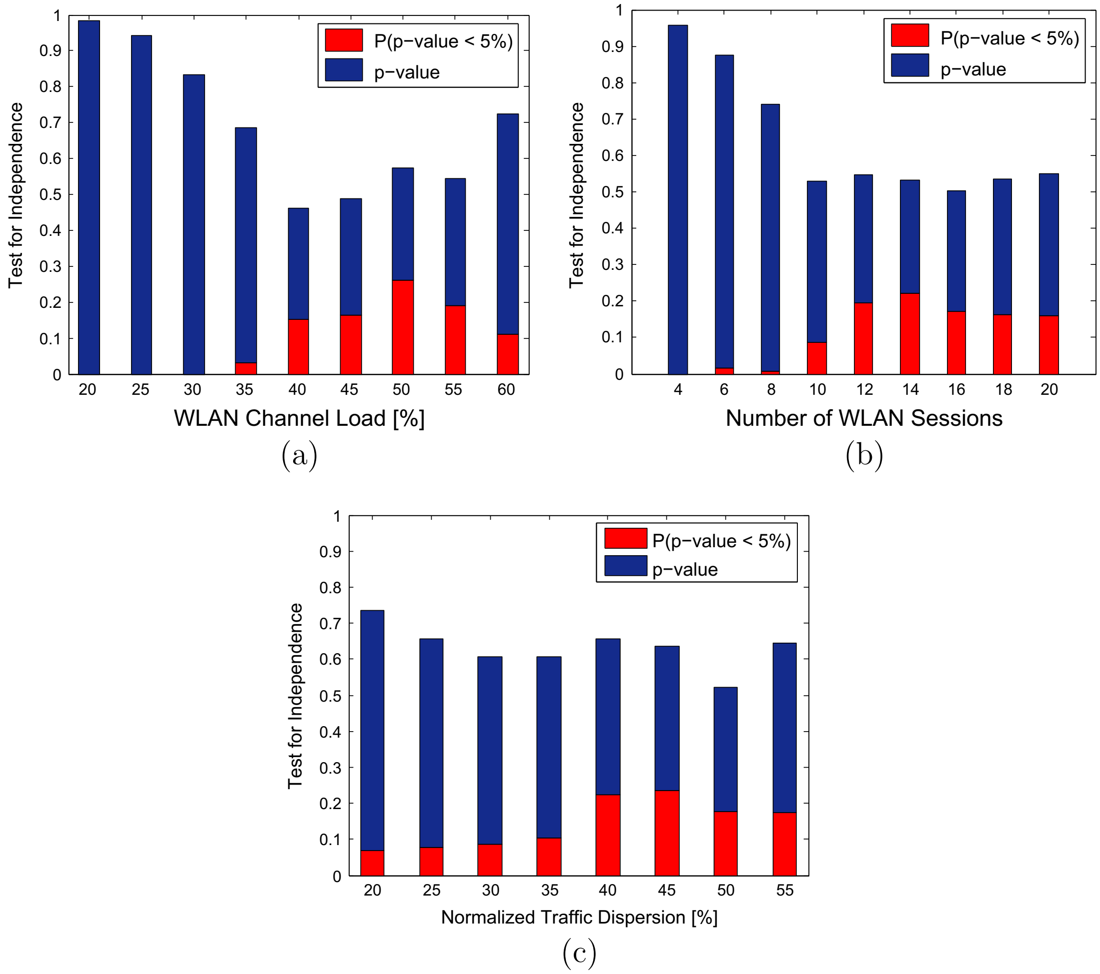


Fig. 20. Industrial-plant WLAN. Test for independence for the idle period process. Average *p*-value and the percentage of *p*-values below the 5% significance level, with respect to (a) the WLAN channel load, (b) the number of WLAN sessions, and (c) the normalized traffic dispersion among the sessions.

observed correlation metric can occur in a sequence of independent samples. We reject the null hypothesis of independence at a significance level of $\alpha = 5\%$. We repeat the process 100 times by randomizing the starting point of the sub-sequences within the corresponding 400-sample intervals, and calculate the average *p*-value and the probability that the null hypothesis of independence is rejected, that is, $P(p\text{-value} < 5\%)$.

Fig. 18 shows the average *p*-value of the independence test for the campus WLAN scenario, along with the percentage of tests with *p*-value below the 5% significance level. Fig. 18(a) shows the *p*-value with respect to the channel load. Under low or moderate load the test shows a high failure rate, suggesting that the successive idle period durations cannot be considered as uncorrelated. The correlation diminishes at high load values. We observe similar trends in Fig. 18(c), where the failure rate is high when the number of sessions is low. Under light load and few sessions only a few flows are intermixed in the AP area. As flows can have very different characteristics in the campus WLAN case, the idle period distribution can change significantly in the event of a flow arrival or departure; this results in correlated successive idle period lengths in parts of the measurement period. At high load, or, as more sessions are active, more flows are transmitted in parallel, and the correlation decreases. Fig. 18(d) shows the results of the independence test as a function of the traffic dispersion. Low level of dispersion typically means several sessions transmitting at an arbitrary point of time, and we experience good independence properties which deteriorate, as the traffic becomes less balanced and a few sessions dominate the traffic.

Since the lag-1 correlation is high in many of the considered cases, it is worth to evaluate how the correlation changes at increased temporal separation. Fig. 18(b) presents the results of the independence test when the autocorrelation is calculated for lag-50 and lag-100. We observe that the correlation remains significant even between temporal separation of 50 samples (lag-50), and it only degrades drastically for lag-100, when the set of active flows is usually changed. This suggests that idle periods need to have a high temporal distance to be safely considered as independent.

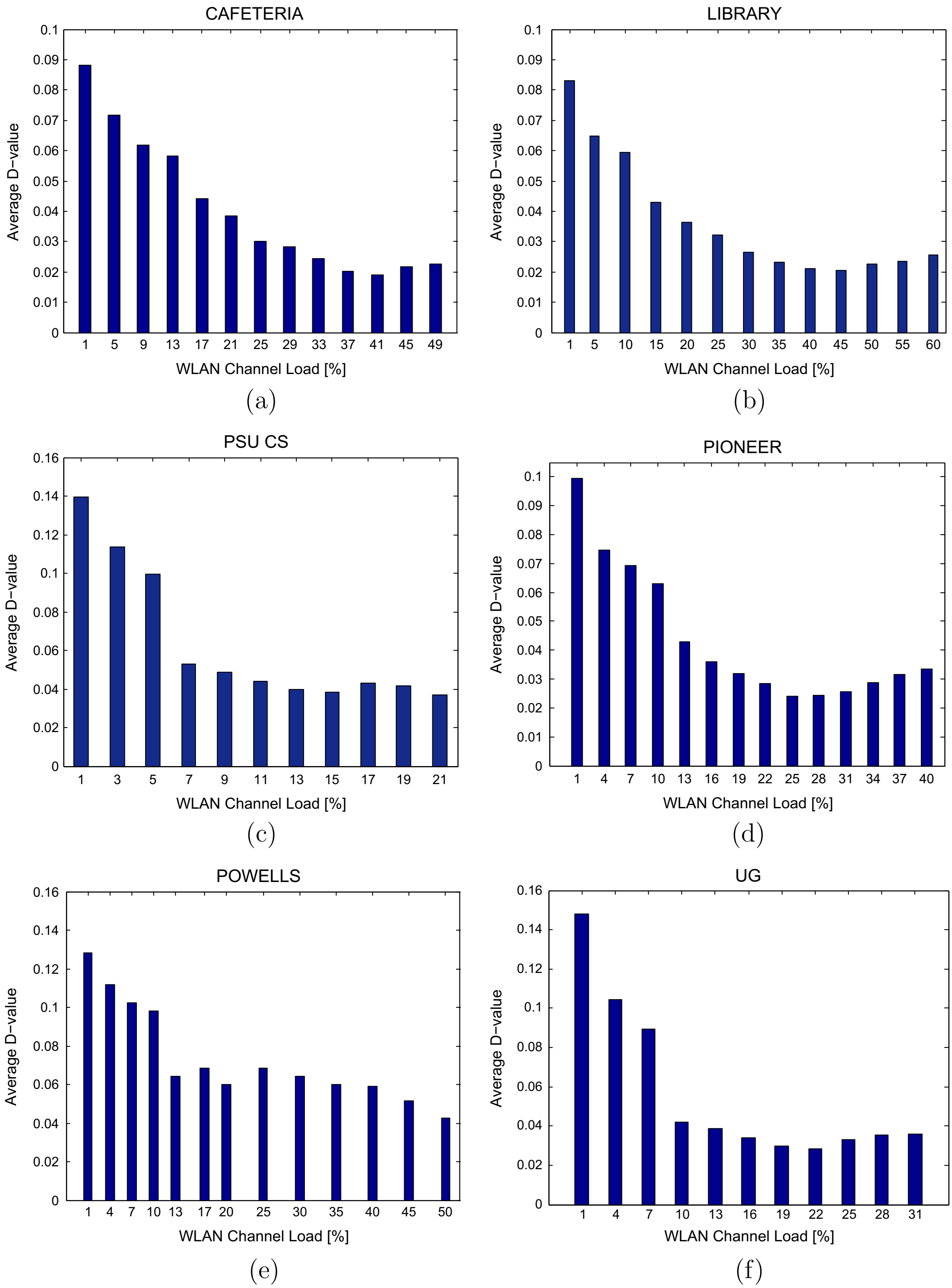


Fig. 21. Aggregate results for fitting accuracy for the considered real WLAN traces

Fig. 19 depicts the *p*-value of the independence test for the conference-hall scenario. As shown in Fig. 19(a) and (b), we observe relatively high *p*-values, that are only marginally affected by the channel load or the number of sessions. The effect of dispersion, as shown in Fig. 19(c), is similar to the one in the campus WLAN scenario (Fig. 18(d)), that is, higher traffic dispersion decreases the *p*-value which, however, remains at a rather high level. According to our simulations, the probability of experiencing high dispersion is low, $(P(\eta > 110) < 0.05)$, and consequently there are very few test results in the high dispersion range. This allows both the *p*-value and the hypothesis rejection probability to be around 0.5 in this range.

In general, the hypothesis on the independence of the consecutive idle periods may hold in the conference-hall scenario. Contrary to the campus WLAN case, the majority of the traffic is from Web flows with identical packet inter-arrival processes. Additionally, flow arrival intensities are higher, leading to a high number of concurrent flows intermixed in the AP area; the above reasons result in an aggregated idle period sequence with low autocorrelation.

Fig. 20 presents the results of the industrial-plant scenario. Here the resulting *p*-values are high, particularly at low and moderate load cases (which also implies low number of sessions in this scenario), as a result of the constant number of sessions and flows with very similar characteristics (Fig. 20(a) and (b)). The *p*-value drops at higher load, when increased contention introduces consecutive back-off idle periods whose statistical properties differ significantly from those of the packet inter-arrival process. The *p*-value increases again at very high load, when most of the idle-periods are back-offs. Finally, Fig. 20(c) indicates that the idle sequence autocorrelation is almost indifferent to the level of traffic dispersion, which is expected, since the traffic dispersion range is very short.

To summarize, the Markovian assumption holds with high probability in the conference-hall scenario and in the industrial-plant WLAN, unless the channel load, and thus the probability of contention are high. In the campus-WLAN scenario, however, the Markovian assumption is only justified under high channel load, when a high number of traffic flows are intermixed. The idle sequence autocorrelation decreases drastically only for a temporal separation in the order of 100 samples. Nevertheless, based on (1) and the experimental results in Fig. 4(a), the corresponding time distance lies below 1 s. This means that the Markovian assumption can be applied for relevant protocol design, e.g. in the case of the coexistence of WLANs and wireless sensor networks with long duty-cycle medium access schemes.

6. Model validation over real WLAN traces

This Section investigates the ability of the proposed 802.11 channel occupancy model to capture the statistical behavior of a set of real WLAN channel usage data. We experiment with a set of real, high time-resolution 802.11 traces, captured with a commercial sniffer device in diverse WLAN environments [32]. The considered data set includes traces from campus wireless networks, namely a campus coffee spot (`CAFETERIA`), a university library building (`LIBRARY`) and a university office building (`PSU_CS`), allowing direct comparison with the results with the synthetic Campus WLAN traffic. The dataset includes, additionally, traces from a public hot-spot (`PIONEER`) and two coffee places (`POWELLS`, `URBAN GRIND (UG)`).

The considered traces provide a high (nano-second) resolution timing information on packet arrivals and frame-in-the-air durations as well as MAC control information including beaconing, and 802.11 control packet transmissions. Thus their level of detail is appropriate for generating channel occupancy statistics, and, consequently, for our objective of WLAN channel usage characterization. For all traces there exists an unfiltered version containing all frames correctly deciphered by the radio capturing device as well as a filtered-by-BSSID ("*pcap*") version limiting the captured trace to the traffic associated with the considered WLAN hot-spot. Note, however, that the higher-layer traffic characteristics are not known.

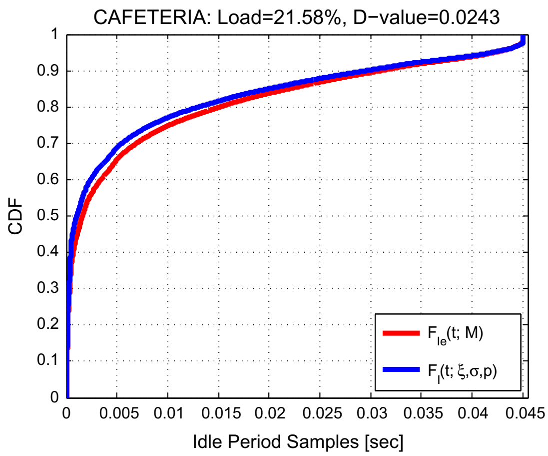
As the traces provide idle period sequences that are in the order of 10^5 samples, we partition them into shorter sub-sequences of $4 \cdot 10^4$ samples each. We perform the model parameterization and fitting validation for all sub-sequences and extract aggregate statistical results, evaluating both the accuracy of the idle-period fitting, as well as the validity of the Markovian assumption.

Fig. 21 depicts the fitting accuracy of the trace set with respect to the average channel load. In the majority of the cases the trends are similar to the ones with the synthetic traces. Table 8 summarizes the fitting accuracy results for the considered traceset. With the exception of the `POWELLS` case the proposed modeling of the idle period duration can effectively capture the statistical behavior of the idle period traces. This is verified by both the low *D*-value (Fig. 21) of the cdf-fitting process, as well as the low fail rate $(P(p\text{-value} < 5\%))$ of the related K–S test. The resulting $P(p\text{-value} < 5\%)$ for the `LIBRARY` and `CAFETERIA` complies with the fail-rate evaluated for the Campus WLAN scenario (Table 7), while for the `PSU_CS` case it is higher, due to the significantly lower WLAN load.

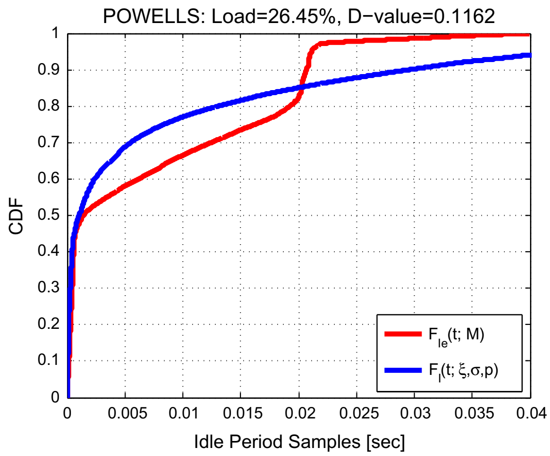
Fig. 22 illustrates a comparative fitting example for sub-sequences taken from the `LIBRARY` and `POWELLS` traces. As shown in Fig. 22(b), the POWELLS trace reflects a rather unexpected, significant nearly-periodic WLAN activity, resulting in a high number of WLAN idle periods around 20msec, which does not allow for an accurate fitting with the generalized Pareto-based white-space distribution.

Table 8
Summary of goodness-of-fit evaluation for the considered real trace-based evaluations.

	D-value	p_{KS}	$C_{p_{KS}}$	$P\{p_{KS} \leqslant 5\%\}$
CAFETERIA	0.0322	0.4932	0.0903	0.0783
LIBRARY	0.0356	0.5332	0.1032	0.0888
PSU_CS	0.0414	0.3903	0.1895	0.1439
PIONEER	0.0399	0.4598	0.1421	0.0912
POWELLS	0.0674	0.0694	0.0637	0.5532
UG	0.0388	0.3802	0.1453	0.0941



(a) CAFETERIA



(b) POWELLS

Fig. 22. Examples of empirical distribution fitting for the LIBRARY and the POWELLS traces.

Table 9
Summary of the test for independence for the idle period process in the considered real trace-based evaluations

	p-Value	$P(p\text{-value} < 5\%)$
CAFETERIA	0.6592	0.0842
LIBRARY	0.8931	0.0702
PSU_CS	0.3931	0.1977
PIONEER	0.5312	0.0951
POWELLS	0.2921	0.3982
UG	0.1932	0.2830

Finally Table 9 summarizes the results of the evaluation of the Markovian assumption. For most of the traces the resulting fail rates of the test for independence and the related p-value are comparable to the ones of the synthetic networking scenarios. As for the fitting accuracy test, the `POWELLS` case with its periodic traffic exhibits a high fail. The fail rate is rather high even for the `UG` rate, showing again that the Markovian assumption has to be handled with care.

7. Conclusions

The modeling of the radio channel occupancy becomes a key issue for the design of future wireless networks sharing common spectrum bands. In this paper we considered the special case of IEEE 802.11 WLANs channel occupancy, since these networks are wide-spread and emerging networks have to adapt their access strategies to WLAN presence.

We addressed the question whether semi-Markovian channel occupancy models already proposed in the literature, but validated only for very limited use cases, can be applied in realistic networking environments. As controlled experiments with large parameter sets are hard to conduct in real testbeds, we performed our study via simulations. For the simulations we selected three networking scenarios, with significantly different traffic workload, the university campus, the conference-hall, and the industrial-plant. We performed detailed study to evaluate whether the proposed WLAN channel occupancy model with heavy-tail idle time distribution and the Markovian assumption on the independence of the consecutive idle times are valid.

Considering the proposed idle time distribution, we can conclude that the accuracy is affected by the traffic mix and the network load. The white space distribution is satisfactory in most of the cases, however, it is not very accurate when the traffic is very heterogeneous and the load is low, or under heavy, nearly periodic traffic. The assumption on uniformly distributed back-off periods necessarily fails under high contention level, for example in cases where the high MAC error rate and consequent retransmissions moves the network to the high load regime. The Markovian assumption holds for many of the considered scenarios, but fails again when the traffic is very heterogeneous and the load is low.

Clearly, both the traffic mix and the networking technology change with time, therefore, it is necessary to discuss the generality of our results. New services introduce flows with new in-flow characteristics, and the weight of the different flow types changes with time. Based on our results we can predict with confidence, that the proposed channel occupancy model will hold in future networking scenarios as well, apart from the cases when the load is low, with some dominant flows. In those cases the in-flow characteristics of the dominant flows determines the idle-time distribution. New networking technologies are expected to increase the transmission rate in general and to use efficient physical layer techniques that will increase spectrum efficiency. As our results show, these changes do not affect the accuracy of the proposed channel occupancy model. Power saving options with duty cycling may introduce periodicity in the channel access of some terminals with low traffic. This will not affect the channel occupancy characteristics if the aggregate load of the AP is not very low.

Acknowledgement

This work was supported by the European STREP Project "HYDROBIONETS" Grant no. 287613 within FP7.

References

[1] Q. Zhao, L. Tong, A. Swami, Y. Chen, Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework, IEEE J. Select. Areas Commun. 25 (3) (2007) 589–600.
[2] X. Li, Q. Zhao, X. Guan, L. Tong, Optimal cognitive access of markovian channels under tight collision constraints, IEEE J. Select. Areas Commun. 29 (4) (2011) 746–756.
[3] K.W. Choi, E. Hossain, Opportunistic access to spectrum holes between packet bursts: a learning-based approach, IEEE Trans. Wirel. Commun. 10 (8) (2011) 2497–2509.
[4] S. Pollin, I. Tan, B. Hodge, C. Chun, A. Bahai, Harmful coexistence between 802.15.4 and 802.11: a measurement-based study, in: Proceedings of ICST International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '08), 2008, pp. 1–6.
[5] S. Geirhofer, J.Z. Sun, L. Tong, B.M. Sadler, Cognitive frequency hopping based on interference prediction: theory and experimental results, ACM SIGMOBILE Mob. Comput. Commun. Rev. 13 (2009) 49–61.
[6] J. Huang, G. Xing, G. Zhou, R. Zhou, Beyond co-existence; exploiting WiFi white space for ZigBee performance assurance, in: Proceedings of IEEE International Conference on Network Protocols, 2010.
[7] A. Balachandran, G.M. Voelker, P. Bahl, P.V. Rangan, Characterizing user behavior and network performance in a public wireless LAN, SIGMETRICS Perform. Eval. Rev. 30 (1) (2002) 195–205.
[8] M. Afanasyev, T. Chen, G.M. Voelker, A.C. Snoeren, Usage patterns in an urban WiFi network, IEEE/ACM Trans. Netw. 18 (5) (2010) 1359–1372.
[9] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, N. Shankaranarayanan, Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots, in: Proceedings of IEEE International Conference on Computer Communications (INFOCOM), 2011, pp. 2921–2929.
[10] G. He, J.C. Hou, W.-P. Chen, T. Hamada, Characterizing individual user behaviors in WLANs, in: Proceedings of ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM '07), 2007, pp. 132–137.
[11] X.G. Meng, S.H.Y. Wong, Y. Yuan, S. Lu, Characterizing flows in large wireless data networks, in: Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom '04), 2004, pp. 174–186.
[12] H. Feng, Y. Shu, O.W. Yang, Research on characterization of wireless LANs traffic, Commun. Nonlin. Sci. Numer. Simul. 16 (8) (2011) 3179–3187.
[13] F. Wamser, R. Pries, D. Staehle, K. Heck, P. Tran-Gia, Traffic characterization of a residential wireless internet access, Telecommunications Systems, vol. 48(1–2), Springer, 2010, pp. 5–17.
[14] J.-K. Lee, J.C. Hou, Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application, in: Proceedings of ACM International Symposium on Mobile ad hoc Networking and Computing (MobiHoc '06), 2006, pp. 85–96.
[15] M. Balazinska, P. Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: Proceedings of ACM International Conference on Mobile Systems, Applications and Services (MobiSys '03), 2003, pp. 303–316.
[16] F. Hernández-Campos, M. Karaliopoulos, M. Papadopouli, H. Shen, Spatio-temporal modeling of traffic workload in a campus WLAN, in: Proceedings of EAI International Workshop on Wireless Internet (WICON '06), 2006.
[17] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE J. Select. Areas Commun. 18 (3) (2000) 535–547.
[18] P. Rathod, O. Dabeer, A. Karandikar, A. Sahoo, Characterizing the exit process of a non-saturated IEEE 802.11 wireless network, in: Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '09), 2009, pp. 249–258.
[19] J. Misic, V. Misic, Characterization of idle periods in IEEE 802.11e networks, in: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '11), 2011, pp. 1004 –1009.
[20] L. Stabellini, Quantifying and Modeling Spectrum Opportunities in a Real Wireless Environment, in: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '10), 2010, pp. 1 –6.
[21] C. Ghosh, S. Pagadarai, D. Agrawal, A. Wyglinski, A framework for statistical wireless spectrum occupancy modeling, IEEE Trans. Wirel. Commun. 9 (1) (2010) 38–44.
[22] S. Geirhofer, L. Tong, B. Sadler, Cognitive radios for dynamic spectrum access – dynamic spectrum access in the time domain: modeling and exploiting white space, IEEE Commun. Magaz. 45 (5) (2007) 66–72, http://dx.doi.org/10.1109/MCOM.2007.358851. http://dx.doi.org/10.1109/MCOM.2007.358851.
[23] S. Geirhofer, L. Tong, B. Sadler, Cognitive medium access: constraining interference based on experimental models, IEEE J. Select. Areas Commun. 26 (1) (2008) 95–105.
[24] I. Glaropoulos, V. Fodor, L. Pescosolido, C. Petrioli, Cognitive WSN transmission control for energy efficiency under WLAN coexistence, in: Proceedings of ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM' 11), 2011, pp. 261 –265.
[25] M. Lagana, I. Glaropoulos, V. Fodor, C. Petrioli, Modeling and estimation of partially observed WLAN activity for cognitive WSNs, in: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC' 12), 2012.
[26] C. Ciflikli, A. Gezer, A.T. Ozsahin, O. Ozkasap, Bittorrent packet traffic features over IPv6 and IPv4, Simul. Model. Pract. Theory 18 (9) (2010) 1214–1224.
[27] D. Bonfiglio, M. Mellia, M. Meo, D. Rossi, Detailed analysis of Skype traffic, IEEE Trans. Multi. 11 (1) (2009) 117–127.
[28] EU FP7 HYDROBIONETS Deliverable D2.2 – Scenario Descriptions and System Requirements (v2), 2013 <http://www.hydrobionets.eu/index.php/deliverables>.
[29] NS-Miracle: Multi-InteRfAce Cross-Layer Extension library for the Network Simulator <http://telecom.dei.unipd.it/pages/read/58/>.
[30] T. Chrysikos, G. Georgopoulos, S. Kotsopoulos, Site-specific validation of ITU indoor path loss model at 2.4 GHz, in: Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks Workshops, (WoWMoM '09), 2009, pp. 1–6.
[31] J.-Y.L. Boudec, Performance Evaluation of Computer and Communication Systems, EPFL Press, 2010.
[32] C. Phillips, S. Singh, Analysis of wlan traffic in the wild, in: Proceedings of the 6th International IFIP-TC6 Conference on Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet, NETWORKING'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 1173–1178.

Ioannis Glaropoulos received his Diploma in Electrical and Computer Engineering with specialization in Electronics & Computers from the School of Engineering at the Aristotle University of Thessaloniki in 2005. In 2008 he received his Master of Science in Electrical Engineering with specialization in Wireless Systems at the Royal Institute of Technology (KTH) in Stockholm. As of October 2008 he is a Ph.D. student at the Laboratory for Communication Networks at the School of Electrical Engineering at KTH. His research interests include dimensioning and routing in wireless sensor networks, cognitive radios and dynamic spectrum access technologies, and stochastic models for spectrum usage in wireless networks.

Alex Vizcaino Luna received his Diploma in Telecommunication Engineering with specialization in Telematics from the School of Engineering at the Universitat Rovira i Virgili in Catalonia in 2010. In 2012 he received his Master of Science in Telecommunication Engineering and Management from the School of Telecommunications and Aerospace Engineering at the Polytechnic University of Catalonia, Spain. In the same year he received his Master of Science in Network Services and Systems from the School of Electrical Engineering at the Royal Institute of Technology (KTH) in Stockholm, Sweden. He is currently with INSA, Spain, working as an ABAP Software Developer for the SAP Healthcare Module.

Viktoria Fodor received her M.Sc. and Ph.D. degrees in computer engineering from the Budapest University of Technology and Economics in 1992 and 1999, respectively. In 1994 and 1995 she was a visiting researcher at Politecnico di Torino and at Boston University, where she conducted research on optical packet switching solutions. In 1998 she was a senior researcher at the Hungarian Telecommunication Company. In 1999 she joined the KTH, Royal Institute of Technology, where she now acts as assistant professor in the Laboratory for Communication Networks. Her current research interests include multimedia and peer-to-peer communication, sensor network and cognitive radio network protocol design and performance analysis.

Maria Papadopouli (Ph.D. Columbia University, 2002) is an Associate Professor in the Department of Computer Science at the University of Crete, a guest Professor at the KTH Royal Institute of Technology in Stockholm, and a Research Associate in the Institute of Computer Science, Foundation for Research and Technology-Hellas. From July 2002 until June 2004, she was a tenure-track Assistant Professor at the University of North Carolina at Chapel Hill. She has co-authored a monograph on Peer-to-Peer Computing for Mobile Networks: Information Discovery and Dissemination (Springer Eds. 2009). In 2004 and 2005, she was awarded with an IBM Faculty Award, and in 2013 with a GOOGLE Faculty Award. In 2013, she received an investigator-driven grant based on scientific excellence from the General Secretariat for Research and Technology in Greece.