802.11 Wireless Access Point Usage Simulation and Anomaly Detection

Anisa Allahdadi · Ricardo Morla

Received: date / Accepted: date

Abstract Despite the growing popularity of 802.11 wireless networks in many public places, wireless users often suffer from connectivity problems due to unstable radio conditions and dynamic user behavior among other reasons. Detecting the anomalous cases and planning to prevent such problems in advance are in the thick of major challenges that network managers encounter. Complication of monitoring such complex networks, that often requires heavy instrumentation of the user devices, makes the anomaly detection analysis even harder. In this paper we exploit 802.11 access point usage data and propose an anomaly detection technique based on Hidden Markov Models and on data that is inexpensive to obtain. We subsequently validate the proposed model by generating a number of network anomalous scenarios in OMNeT++/INET network simulator and evaluate the detection method against baseline approach (Raw_Data and PCA) results. The experimental outcomes show the superiority of the proposed HMM model in detection precision and sensitivity.

Keywords 802.11 Access Points · Network Management · Anomaly Detection · Hidden Markov Models · Network Simulation · OMNet++/INET

1 Introduction

In recent years, IEEE 802.11 wireless networks has emerged as a promising technology for wireless access by mobile devices in many public places from enterprises and universities to urban areas. The flourishing popularity and ease of access to these networks has led to heavy utilization and congestion circumstances. On the other hand, interferences caused by broadcast nature of wireless links along with the other radio waves in the same frequency range normally result in poor performance as the packet transmission fails or requires several retransmission attempts. Furthermore, dynamic traffic loads and evolving nature of users movement and association to APs often induce connectivity problems in large-scale 802.11 deployments. Generally speaking, at any given moment 802.11 APs or users are likely to come across problems threatening the connection quality. Thus the question of performance becomes increasingly important as new applications demand sufficient bandwidth and reliable medium access.

The task of detecting and fixing the root causes of performance degradation is among the major challenges of network administration team. Across the infrastructure, there are various types of anomalous situations caused by users or APs, and automatic detection of these anomalies is of great importance for future mitigation plans. Highly utilized medium, overloaded APs, failed or crashed APs, persistent interference between adjacent APs, RF effects and authentication failure are examples of such anomalies. However, due to the timely and costly limitations of constantly monitoring the entire wireless field by sensors and sniffers, obtaining reliable ground truth becomes more and more challenging.

In such circumstances, network simulations are imminent solutions to obtain a setup that is computationally tractable by making some simplifying assumptions. In the research community, many wireless networks are evaluated using discrete event simulators like OMNet++ [1] [2] [3]. Although having worked with other
simulation frameworks such as NS3 and OPNET, we found OMNeT++/INET the most appropriate wireless network simulators for our research purposes. Besides the well-structured framework and user-friendly IDE that facilitate analysis and results acquisition, OMNeT++/INET provides an adequate set of modules supporting physical and radio models for 802.11 that perfectly meet our requirements for this project.

In our previous papers [4][5], we utilized RADIUS authentication log data collected at the hotspot of the Faculty of Engineering of the University of Porto (FEUP). The trace data consists of the daily summary of the connections between hundreds of APs and their corresponding wireless stations. In [6], we deployed a real testbed in small scale with one AP and 6 STAs using FreeRADIUS server, and generated a number of anomalies in a controlled environment for experimental purposes. In this work, we simulate a mini WLAN with 2 APs and 16 STAs and set up several anomalous cases, including the previous ones in [6] and some new anomalies. We further evaluate our HMM model formerly proposed in [5] and [6] with the simulation results.

The key steps of the work include: 1) Conduct 802.11 wireless network simulation in OMNeT++/INET to resemble normal and anomalous scenarios. 2) Reiterate the simulations with different seeds to provide miscellaneous replicates. 3) Extract the simulation results, and convert them to sequential data. 4) Build HMM models from the prepared data set and apply the proposed anomaly detection algorithm. 5) Calculate the detection rate and evaluate the model based on the normal and anomalous periods.

The rest of the paper proceeds as follows. In section 2, the related work and the most recent researches relevant to the current work are presented. In section 3, the anomaly detection methodology is described briefly and the previous studies on this subject are indicated. Section 4 deals with the network simulation setup and focus on the common key properties of the accomplished simulations. In section 5 the simulated scenarios are described and the experimental results are analyzed. In section 8, the main conclusions are provided and the prominent direction of future work is disclosed.

2 Related Work

2.1 Anomalous Patterns Detection

In the most recent studies concerning 802.11 wireless networks, there exist several analysis on connectivity and performance issues for facilitating the network management tasks. In connection to this, a number of articles investigate the overloaded networks, faulty APs, impact of interference in chaotic 802.11 deployments and similar anomalous cases.

Having explored the network under high medium utilization conditions, authors in [7] show that in the over-loaded networks, stations only maintain a short association period with an AP, and repeated association and re-association attempts are common phenomena even in the absence of client mobility. Their analysis demonstrates that stations throughput suffers drastically from the unnecessary handoffs, leading to suboptimal network performance.

In another direction of work in [8], the authors present a number of algorithms that detect failed APs by analysis of AP usage logs. The main assumption in their algorithm is that the longer the time an AP does not register events, the greater the probability that particular AP is faulty (crashed/halted).

In relation to interference detection in WLANs, authors in [9] propose methods including intelligent frequency allocation across APs, load balancing of user affiliations across APs and AP adaptive power control for interference mitigation in dense 802.11 deployments. Furthermore, the authors in [10] studied the impact of RF interference on 802.11 networks from devices like Zigbee and cordless phones that crowd the 2.4GHz ISM band to devices like wireless camera jammers and non-compliant 802.11 devices that disrupt 802.11 operations. They affirm through practice that moving to a different channel is more effective in coping with interference than changing 802.11 operational parameters such as CCA (clear channel assessment).

In [11], a usage pattern called ”abrupt ending” is explored in FEUP data set that concerns the disassociation of a large number of wireless sessions in the same AP within a one second window, or in a nutshell ”simultaneous session ending”. The authors introduce some anomalous patterns that might be in correlation with the occurrence of this phenomena. For instance, AP halt/crash, AP overload, persistence interference and intermittent connectivity. The analysis of the anomaly-related patterns performed in this research, inspired our work to regenerate similar anomalies in network simulator in addition to the real Testbed that was already done in our previous work [6]. The principal goal of the simulation and the real Testbed experiments is to evaluate the HMM anomaly detection methodologies proposed in the current work as well as our former studies [5][6].

2.2 Wireless Network Simulation

There are numerous efforts in the literature that tried to exploit simulation as an effective tool to setup a
computationally tractable network. Wireless network simulation is used for various objectives from assessment and validation of models to obtain synthesized data and parameterized metrics. In [12] the authors employed simulation to generate synthetic traffic and validate their proposed model of traffic workload in a campus WLAN. As another example, the researchers in [13] propose a framework to integrate the infrastructure mode and ad hoc mode and they implement the framework in NS2. They used simulation to show the higher performance of their proposed model compared to the traditional wireless LAN. In a rather relevant work to ours, the performance of IEEE 802.11 wireless networks is evaluated using OPNET Modeler [14]. The simulated network in infrastructure mode for one AP and 12 stations investigates the performance of pure 802.11g network over a network that uses both 802.11g and 802.11b clients.

In relation to OMNet++ and its simulation models, a number of articles work on validating the reliability and accuracy of OMNeT++. For example in [15] the authors perform a measurement study of wireless networks in a highly controlled environment to validate the IEEE 8021.11g model of OMNeT++. They used metrics like throughput, delay and packet inter-transmission to compare the measurement results to identical simulations. They show that the simulation results match the measurements well in most cases. Furthermore in [16] the reliability of OMNeT++ is assessed for wireless DoS attacks by comparing the simulation results to the real 802.11 testbed. In this case throughput, end-to-end delay, and packet lost ratio are considered as performance measures. The authors confirm the accuracy of the simulation results in wireless DoS domain.

However, there exist few efforts in the literature that conduct simulation of WLANs in OMNeT++ and concern about performance and quality of service (QoS). For example in [2] the performance of the TCP protocol for audio and video transmission is evaluated using OMNeT++ simulation. In another direction of work in [3] an overview of the IEEE 802.11b model is simulated in OMNeT++ and an example network consisting of a mobile station moving through a series of APs is used to analyze the handover behavior of the model. To the best of our knowledge the simulation of aforementioned anomalous patterns in WLAN infrastructure mode has never been done before.

2.3 HMM Applications in Network Analysis

In wireless networking, HMMs are employed to address various aspects of network measurement and analysis. Hierarchical and Hidden Markov based techniques are analyzed in [17] to model 802.11b MAC-to-MAC channel behavior in terms of bit error and packet loss. The authors employed two random variables in packet loss process, inter-arrival-rate and burst-length of packet loss, and applied the traditional two-state Markov chain. The results demonstrates that two-state Markov chain provides an adequate model for the 802.11b MAC-to-MAC packet loss process.

In a more recent line of work in [18] a multilevel approach involving HMMs and Mixtures of Multivariate Bernoullis (MMB) is proposed to model the long and short time scale behavior of wireless sensor network links, that is, the binary sequence of packet receptions (1s) and losses (0s) in the link. In this approach, HMM is applied to model the long-term evolution of the trace, and the short-term evolution is modeled within the states by HMM or M Mb. The notion of multilevel HMM, or higher dimensional HMM, is an impressive concept regarding to our own work, and we intend to make use of this approach to improve our HMM variations for anomalous pattern recognition in the future work.

In another related work, HMMs are applied for modeling and prediction of user movement in wireless networks to address QoS issues [19]. User movement from an AP to an adjacent AP is modeled using a second-order HMM. Although the authors demonstrated the necessity of using HMM instead of Markov chain model, the proposed model is only practical for small wireless networks with a few number of APs, not widespread WLANs.

As the above literatures indicate, HMM related studies in wireless network management are rarely used specifically in performance anomaly detection.

3 Data Set Description

In our previous papers we utilized RADIUS authentication log data which contains session records of wireless stations connecting to APs. A preliminary analysis on the raw data yields a sequential data set summarizing APs association history. In the current simulation we create a similar data set with the exact same features to be synchronized with the previous HMM modeling. The definition of the main features along with a brief explanation on the feature selection process is presented in the following paragraphs:

3.1 Data Features

Data features are categorized in two main classes: Density Attributes and Usage Attributes. Density Attributes demonstrate how crowded is the place in terms of active
attendant users, and the *Usage Attributes* disclose the volume of the sent and received traffics by the present users. The former attributes mainly characterize the association population and durability, and the later ones reveal the total bandwidth throughput regardless of how populous is the place and it is more relevant to the applications utilized by the current mobile users.

3.1.1 Density Attributes

*User Count*: the number of unique users observed in a specific location (indicated by an AP) in a time-slot.

*Session Count*: the total population of active sessions during a time-slot regardless of the owner user. This attribute reveals the number of attempts made by the the congregation of the present users to associate to the current AP.

*Connection Duration*: the total duration of association time of all the current users. This attribute is an indicator of the overall connection persistence. The utmost amount of this features is achieved when there is no evidence of disassociation in the ongoing active sessions during a time-slot.

3.1.2 Usage Attributes

*Input Data in Octets*: the number of octets transmitted from the client. This attribute briefly refers to the number of bytes uploaded by the wireless user.

*Output Data in Octets*: the number of octets received by the client. This attribute shortly refers to the number of bytes downloaded by the wireless user.

*Input Data in Packets*: the number of packets transmitted from the client. This attribute is similar to the above *Input-Octet*, just to be measured in packets.

*Output Data in Packets*: the number of packets received by the client. This attribute is similar to the above *Output-Octet*, just to be measured in packets.

3.2 Feature Selection

For subsequent analysis, we favor using less features than the entire set of attributes introduced earlier. For this purpose, we applied Principal Component Analysis (PCA) technique to find the combination of the variables which best explain the phenomena and contain the greatest part of the entire information. In the current experiment the first three principal components bring the cumulative proportion of variance to over 99%. More detail explanation on the correlation of data features with themselves and with the principal components are provided in our previous work [6].

4 Anomaly Detection in AP Usage Data

We use Hidden Markov Models as a time-variant modeling approach in which conditional probabilities of events are determined based on their history. In this section we explain how the HMM models are initialized, optimized, and utilized for anomaly detection. We also present the automatic threshold recognition procedure using histograms.

4.1 Hidden Markov Models

Rabiner and Juang [21] presented a comprehensive tutorial on HMM which provides a profound understanding of the basic blocks of HMM. HMM symbolizes a doubly stochastic process with a set of observable states and a series of hidden states which can only be observed through the observable set of stochastic process.

The formal definition of a n-state continuous HMM notation is determined as follows:

- A set of hidden states $S = \{s_i\}, 1 \leq i \leq n$
- State transition probability distribution (aks. transition matrix). $A = \{a_{i,j}\}, 1 \leq i,j \leq n, a_{i,j} = P(s_j \text{ at } t+1|s_i \text{ at } t)$
- Observation probability distributions, typically from a normal distribution (aks. emission matrix) $B = \{b_i(o_t)\}, b_i = P(o_t \text{ at } t|s_i \text{ at } t), 1 \leq i \leq n$
  
  $b_i(o_t) = \frac{\exp\left[-\frac{1}{2}(o_t - \mu_i)\Sigma_i^{-1}(o_t - \mu_i)\right]}{(2\pi)^{D/2}|\Sigma_i|^{1/2}}$ where $D$ refers to the dimensionality of the observation space
- Initial state distribution $\pi = \{\pi_i\}, 1 \leq i \leq n$, $\pi_i = P(s_i \text{ at } t = 1)$
- $n = \text{number of hidden states}$

The set $\lambda = (A,B,\pi)$ completely defines an HMM. HMM serves three principal problems of interest listed below [20]:

Problem 1 – Given the observation sequence $O = o_1, o_2, ..., o_T$ and the model $\lambda = (A,B,\pi)$, how we compute $P(O|\lambda)$, the probability of the observation sequence (the Forward-Backward algorithm).

Problem 2 – Given the observation sequence $O = o_1, o_2, ..., o_T$, how we choose a state sequence $S = s_1, s_2, ..., s_T$, which is optimal in some meaningful sense (the Viterbi algorithm).
Problem 3 – How we adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$ (the Baum-Welch algorithm).

In our model we consider fully connected HMMs (ergodic model) and continuous observations with Gaussian distribution and 3 hidden states initialized randomly. We tested the models with 2 and 4 states as well, but the best practice for this problem belongs to the HMMs with 3 states. The HMMs with 2 states are very simple to capture the diverse characteristics of the locations (APs), while there is not enough variety in day-long time series for 4 or higher number of states.

Regarding the initialization process, the observations contain multivariate Gaussian distribution and each component of the mean vector is uniformly drawn between $\mu - 3\sigma$ and $\mu + 3\sigma$. Furthermore the initial covariance matrix is diagonal and each initial variance is uniformly drawn between $\frac{1}{2} \sigma^2$ and $3 \sigma^2$. The initial probability matrix ($\pi$) and the transition matrix ($A$) are uniformly drawn. The initial HMM is then optimized by the Baum-Welch algorithm with the cut off likelihood value of $1e^{-6}$ or the maximum number of iterations set to 20.

Ensuing the optimization process, the physical meaning of the hidden states are more discernible. The values of the principal components in each state shows the tendency of the states to the usage or density attributes. For example a hidden state with a high value for the second principal component shows a more populated case in terms of the number of users. In the future work where the concern is more on modeling the anomalous patterns we utilize the interpretation of the hidden states to relate them to the physical conditions of the locations.

4.2 Likelihood Series

The likelihood of HMMs is basically the first key problem of HMMs stated earlier: the probability of an observation sequence given the model parameters. The following equation shows how the likelihood value of HMM model of $\lambda$ is calculated.

$$P(O|\lambda) = \sum_{a(t)S} P(O|S, \lambda)P(S|\lambda)$$

$$= \sum_{s_1,s_2,\ldots,s_T} \pi_{s_1}b_{s_1}(O_1)a_{s_1,s_2}b_{s_2}(O_2)a_{s_2,s_3}\ldots a_{s_{T-1},s_T}b_{s_T}(O_T)$$

Due to the vanishingly small likelihood probabilities produced in long time-series, the logarithmic value is calculated. However, having built the HMM model for the normal scenario, the log-likelihood values of a single data point or a series of sequential data points is evaluated based on the above equation. The unexpected low log-likelihood values show the divergence from the normal model and are typically indicative of anomalies.

Figure 1 demonstrates the log-likelihood series of an example anomaly generated via simulation. The anomalous case is related to heavy usage of wireless users of a selected AP.

The potential root causes of low log-likelihood in terms of HMM characteristics are analyzed in our previous papers [5, 6].

4.3 Threshold Detection

To detect the anomalous points in the likelihood series, some machine learning techniques need to be applied to determine the anomaly threshold automatically. This threshold defines a boundary for the likelihood series where the lower values belong to the anomalous set. As many anomaly detection algorithms presume, outliers are the minority group not following the common pattern of the majorities. Accordingly we look for the extreme data points (outliers) with the lowest likelihood values. To this end an univariate histogram is constructed and the relative frequency (height of the histogram) is computed. The frequency of samples falling into each bin is used as an estimate of the density. We assume the samples with the highest density (mode) are the normal data points, and accordingly the bins containing the lowest frequencies and farther from the mode bin are the outliers. As a rule of thumb we start checking from the mode bin and stop when a bin with lower frequency than a quarter of mode is observed. Like any other change detection algorithm ours as well produce false
alarms, however in all the performed experiments of this work the false positive ratio is insignificant. We use the same algorithm to detect the outliers or anomalies in Raw_Data and PCA for the purpose of comparison. However, as Raw_Data contains seven features and PCA has 3 components, we conduct the algorithm on each single feature and aggregate the detected points as the final outcome. For example for the likelihood series of s₁s₂...s₄₀, the algorithm detects s₂ and s₄ for the first feature and s₄ and s₁₅ for the third feature and for the rest of the features no anomaly is detected. In this case the final anomalous set contains {s₂, s₄, s₁₅}.

The marked data points on the likelihood series of heavy usage example in Figure 1 are achieved applying this algorithm.

5 Experimental Setup

In order to evaluate the proposed strategy, we perform an extensive set of simulations using OMNeT++ simulator and INET framework. OMNeT++ is a C++-based discrete event simulator (DES) for modeling communication networks, multiprocessors and other distributed or parallel systems [21]. It has a generic architecture and is used in various problem domains including the modeling of wired and wireless communication networks. One of the major network simulation model frameworks for OMNeT++ is the INET Framework [22] that provides detailed protocol models for TCP, IPv4, IPv6, Ethernet, Ieee802.11b/g, MPLS, OSPFv4, and several other protocols. We used OMNeT++ along with INET Framework to simulate the IEEE 802.11 WLAN in infrastructure mode which is the network model in many public places including university campuses.

In a discrete event simulator, as well as the OMNeT++, events take place at discrete instances in time, and they take zero time to happen. It is assumed that nothing important happens between two consecutive events. Thus the simulation time is relevant to the order of events in the events’ queue, and it could take more than the real CPU time or less than it based on the number of nodes, amount of traffic transferred, and other details of the network. In our example, with the current number of nodes and traffic plan, 100s of simulation time takes around 5min of CPU time. Our HMM approach operates on 40 consecutive time slots of 15s simulation time each.

5.1 Normal Scenario

Figure 2 shows the initial picture of a normal scenario, the location of the access points (APs), wireless stations (STAs), and the servers. In the normal scenario, there are 2 APs and 16 STAs initially associated to one AP depending on their location. During the simulation STAs, based on their mobility models, are handed over to the other AP when moving around the simulation ground. Furthermore, according to the defined traffic plans in section 5.3, each node sends and receives packets to the existing servers.

5.2 Mobility Models of the Wireless Stations

The APs are stationary and the wireless nodes follow different mobility patterns. In the current experiment, the mobility models of the nodes are selected in a way so that the usage behavior of two typical places in a campus could be emulated. The mobile nodes initially connected to the first AP (AP1) follow Linear Mobility pattern which is configured with speed, angle and acceleration parameters. The mobile nodes move to random destinations with the specified parameters and when they hit a wall they reflect off the wall at the same defined angle. These nodes connect to the other AP (AP2) besides their own AP (AP1), and sometimes lose the connection when they move to blind spots. This pattern is selected to symbolize the nodes with some degree of freedom but within a limited space like administrative offices.

The nodes connected to the second AP (AP2) follow the Mass Mobility model, and accordingly move within the room. This pattern of mobility is intended to represent places like classroom or library in which users do not leave the place frequently, but still have some motions in the place.
5.3 Traffic Generation

As it is shown in Figure 2 there are three main servers wire connected to the Ethernet switch: srvHostVideo, srvHostFTP, and srvHostEcho. The traffic transferred between wireless stations and the servers (through APs) is considered to be User Datagram Protocol (UDP). The video server (srvHostVideo) sends UDP packets with the normal message length of $N(600B, 150B)$ to the clients of AP2, resembling the video traffic to those stations. The FTP server (srvHostFTP) is to receive the FTP uploads by the clients of AP1 with normal message length of $N(500B, 100B)$. In addition to exclusively downloading or uploading, the other server (srvHostEcho) is in charge of sending and receiving traffics to all the users. This traffic pattern represents the common act of email checking and web browsing by all the wireless nodes. The echo packets length are configured to be smaller than the previous ones, $N(200B, 50B)$, displaying lighter traffic transmission. In AP Overload anomalous scenario one more server is added to take care of heavy channel utilization (srvHostBurst).

6 Anomalies and Experimental Results

In this section we describe the various experiments conducted and demonstrate the HMM anomaly detection results. We explore a set of anomalous scenarios and discuss different cases of each one. Then we present the HMM results and compare them to baseline approach (Raw Data and PCA) results for evaluation purposes.

6.1 AP Shutdown/Halt

When there is no session recorded in RADIUS accounting table for a period of time and a given AP, it is likely that the AP has stopped working - possibly due to a technical problem or power failure. In our simulation, we
Fig. 5 Precision and recall boxplot of Raw Data, PCA and HMM belong to noise scenario. Left: -90dBm, middle: -95dBm, right: -100dBm.

Fig. 6 Log-likelihood series of noise scenario for -95dBm and -105dBm noise level.

reproduced this anomaly by turning off the AP power deliberately during the halt_period for some time_slots. We performed three variations of this scenario as follows:

- halt_period > time_slot.
- halt_period = time_slot.
- halt_period < time_slot.

For each experiment we repeated the simulation 10 times with different seeds to be able to examine the model on rather different sets.

Figure 3 demonstrates the HMM likelihood series and the detected anomalies for the three different variants of this scenario as well as the case without anomalies. The valley shapes in these images show the sudden drops of the likelihood values during the anomalous periods, and the marked points are the anomalies detected by the aforementioned Threshold Detection algorithm. Note that the precision of this algorithm is not ideal and there are some points where likelihood falls are not detected by this algorithm. This figure shows that HMM detects very short shutdown periods that last exactly one time-slot (Figure 3b) or less than one time-slot (Figures 3c). The wider valleys in Figure 3a are related to longer halt_periods.

Figure 4 shows the boxplot diagram of the anomaly detection’s precision and recall computed for Raw Data, PCA and HMM models. In these experiments HMM achieves higher precision values producing smaller false positives compared to the other two approaches which is an eminent aspect in anomaly detection. However, the recall outcomes show that in this scenario HMM is not able to recognize all the anomalous cases while the other two methods are. Note that this type of anomaly is not very difficult to detect just by looking at Raw Data as there is a visible change in data set features when the power is gone and no connection is recorded.

6.2 Noise

Thermal noise, cosmic background noise, and other random fluctuations of the electromagnetic field affect the quality of the communication channel. This kind of noise doesn’t come from a particular source, nor propagate through space [23]. If the noise level is too high, the signal strength will degrad and the performance will decrease.

In the current experiment we change the level of noise power by adjusting the value of IsotropicBack-
802.11 Wireless Access Point Usage Simulation and Anomaly Detection

Fig. 7 Precision and recall boxplot of Raw Data, PCA and HMM belong to AP overload scenario. Left: burst_duration < sleep_duration, middle: burst_duration = sleep_duration, right: burst_duration > sleep_duration.

Fig. 8 The log-likelihood series and detected anomalies of AP overload scenario.

groundNoise parameter in the simulator. The default value of this parameter is set to -110dBm which is the minimum noise level in Wi-Fi networks 802.11 variants. We gradually increased the noise power to -90dBm and recorded the simulation results repeated 10 times for each experiment. According to the study in [24], the average noise level in a busy university campus had a stable value at around -94 dBm.

As the noise power increases, the packets are less likely to be received at the STAs. Therefore two data features are affected directly by the alteration of noise level: OutputOctets and OutputPackets. For this reason the Raw_Data detector is expected to produce satisfactory detection results. However, as Figure 5 shows, HMMs in all the experiments present higher precision values rather than Raw_Data and PCA. However in the third experiment not all the anomalous points are detected by HMM in some simulation runs. As the noise power decreases (higher negative value), it gets more difficult to detect the anomalous periods as the data becomes closer to the normal model.

Furthermore, the noise power of -105dBm is also simulated which is very similar to the normal model (-110dBm). The likelihood series of the HMM model along with the detected anomalies for -95dBm and -105dBm examples are shown in Figure 6. Note that the anomalous period is during the first 10 time-slots.

6.3 AP Overload

In this anomalous case, the excessive channel utilization occurs that could be the consequence of excessive download or upload of a number of wireless users. In such circumstances, even with the high signal strength the clients could get disconnected from the current AP frequently. In this experiment we simulated AP heavy usage caused by all of the users of the second AP in three different situations as following:

- burst_duration < sleep_duration.
• burst\_duration = sleep\_duration.
• burst\_duration > sleep\_duration.

Burst server (srvHostBurst) sends UDP packets to the given IP addresses in bursts during the burst\_duration which resembles the heavy downloads of those users. In the time of sleep\_duration the bursts stop and the channel utilization is back to normal.

Like the previous scenario, each experiment is performed 10 times with different random seeds. Figure 7 displays the boxplot diagram of the precision and recall results of Raw\_Data, PCA and HMM models. The low precision ratios of Raw\_Data and PCA show that this type of anomaly is not straightforward to detect directly from the raw data. The HMM results both in precision and recall outperform the other two methods with a discernible distance.

Figure 8 displays the log-likelihood series of three types of burst\_duration and sleep\_duration obtained for AP overload scenario. As it is shown in these figures, during the burst period the log-likelihood value drops drastically and in the sleep period it increases again to the normal level. The longer the burst period the wider is the valley shape in the log-likelihood series, and HMM effectively detects heavy utilization periods in all different cases.

6.4 Flash Crowd

In wireless networks an unexpected surge of traffic occurs mostly due to the beginning or ending of an event when the majority of the wireless users abruptly enter or leave a place and consequently associate to or disassociate from the current AP. Such incidents are not necessarily an anomaly in terms of performance or connectivity issues, but could be considered more as a sudden change to a routine network. To see whether the HMM model is able to detect such alterations in the normal usage pattern, we simulate this example in two experiments:

• Arrival: simultaneous association of 7 new nodes to the current AP.
• Departure: simultaneous disassociation of 7 existing nodes from the current AP.

As it is illustrated in Figure 9, HMM easily outperforms the Raw\_Data and PCA results in both Arrival and Departure cases.

7 Conclusions and Future Work

Intelligent detection of anomalies in 802.11 networks from the analysis of the collected AP usage data is of great importance to network managers. It facilitates their everyday administration workload as well as assisting them in maintenance of these networks and providing mitigation plans for the future.

The key contributions of this work are HMM modeling and threshold detection technique for anomaly detection in 802.11 wireless networks evaluated under various anomalous scenarios. The precision and recall ratios of the experimental cases are computed and compared to baseline approach (Raw\_Data and PCA). The experimental results show that HMM models are ca-
able of detecting a great portion of anomalies while producing a trivial portion of false positive.

In future work we intend to propose an unsupervised anomaly detection algorithm that automatically distinguish anomalous periods from normal periods and from each other. Furthermore, we plan to proceed to more complex HMMs for learning and characterizing various anomaly-related patterns.

Acknowledgements This work is financed by the ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/FCT/0145-FEDER-006961, and by National Funds through the European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961. The first author is also sponsored by FCT grant SFRH/BD/99714/2014.

References

1. Colesanti, U. M., Crociani, C., & Vitaletti, A. (2007). On the accuracy of omnet++ in the wireless sensor networks domain: simulation vs. testbed. In Proceedings of the 4th ACM workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks (pp. 25–31).: ACM.

2. Qashi, R., Bogdan, M., & Haenssgen, K. (2011). Analysis of packet throughput and delay in ieee 802.11 wans with udp traffic. In International Conference on Mobile Communications, Networking and Applications (MobiCONA). Proceedings (pp. M48).: Global Science and Technology Forum.

3. Woon, S., Wu, E., & Sekercioglu, A. (2003). A simulation model of ieee802.11b for performance analysis of wireless lan protocols. In Australian Telecommunications, Networks and Applications Conference (ATNAC), volume 162.

4. Allahdadi, A., Morla, R., Aguiar, A., & Cardoso, J. S. (2013). Predicting short 802.11 sessions from radius usage data. In Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on (pp. 1–8).: IEEE.

5. Allahdadi, A., Morla, R., & Cardoso, J. S. (2014). Outlier detection in 802.11 wireless access points using hidden markov models. In Wireless and Mobile Networking Conference (WMNC), 2014 7th IFIP (pp. 1–8).: IEEE.

6. Allahdadi, A. & Morla, R. (2017). Anomaly detection and modeling in 802.11 wireless networks. arXiv preprint arXiv:1707.00948.

7. Raghavendra, R., Belding, E. M., Papagiannaki, K., & Almeroth, K. C. (2010). Unwanted link layer traffic in large ieee 802.11 wireless networks. IEEE Transactions on Mobile Computing, 9(9), 1212–1225.

8. Pan, H. & Keshav, S. (2006). Detection and repair of faulty access points. In Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE, volume 1 (pp. 532–538).: IEEE.

9. Broustis, I., Papagiannaki, K., Krishnamurthy, S. V., Faloutsos, M., & Mhatre, V. P. (2010). Measurement-driven guidelines for 802.11 wlan design. IEEE/ACM Transactions on Networking (TON), 18(3), 722–735.

10. Gummadi, R., Wetherall, D., Greenstein, B., & Shen, S. (2007). Understanding and mitigating the impact of rf interference on 802.11 networks. ACM SIGCOMM Computer Communication Review, 37(4), 385–396.

11. Massa, D. & Morla, R. (2013). Abrupt ending of 802.11 ap connections. In Computers and Communications (ISCC), 2013 IEEE Symposium on (pp. 000348–000353).: IEEE.

12. Hernández-Campos, F., Karaliopoulos, M., Papadopouli, M., & Shen, H. (2006). Spatio-temporal modeling of traffic workload in a campus wlan. In Proceedings of the 2nd annual international workshop on Wireless internet (pp.1).: ACM.

13. Chen, J., Chan, S.-H., & Liew, S.-C. (2003). Mixed-mode wlan: the integration of ad hoc mode with wireless lan infrastructure. In Global Telecommunications Conference, 2003. GLOBECOM’03. IEEE, volume 1 (pp. 231–235).: IEEE.

14. Kulgachev, V. & Jasani, H. (2010). 802.11 networks performance evaluation using opnet. In Proceedings of the 2010 ACM conference on Information technology education (pp.149–152).: ACM.

15. Bredel, M. & Bergner, M. (2009). On the accuracy of ieee 802.11 g wireless lan simulations using omnet++. In Proceedings of the 2nd International Conference on Simulation Tools and Techniques (pp.81).: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

16. Malekzadeh, M., Ghan, A. A. A., Subramaniam, S., & Desa, J. M. (2011). Validating reliability of omnet++ in wireless networks dos attacks: simulation vs. testbed. IJ Network Security, 13(1), 13–21.

17. Khayam, S. A. & Radha, H. (2003). Markov-based modeling of wireless local area networks. In Proceedings of the 6th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems (pp.100–107).: ACM.

18. Kamthe, A., Carreira-Perpinán, M. A., & Cerpa, A. E. (2009). M&m: multi-level markov model for wireless link simulations. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (pp. 57–70).: ACM.
19. Prasad, P. S. & Agrawal, P. (2010). Movement prediction in wireless networks using mobility traces. In Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE (pp. 1–5).: IEEE.

20. Rabiner, L. & Juang, B.-H. (1986). An introduction to hidden markov models. ASSP Magazine, IEEE, 3(1), 4–16.

21. OMNeT++ Home Page. http://www.omnetpp.org. Accessed in March 2017.

22. INET Framework Home Page. http://www.inet.omnetpp.org. Accessed in March 2017.

23. INET Framework for OMNeT++ Manual. https://omnetpp.org/doc/inet/api-current/inet-manual-draft.pdf. Accessed in March 2017.

24. Fu, B., Bernth, G., Steichen, B., & Weber, S. (2008). Wireless background noise in the Wi-Fi spectrum. In Wireless Communications, Networking and Mobile Computing, 2008. WiCOM’08. 4th International Conference on (pp. 1–7).: IEEE.

Anisa Allahdadi received the B.Sc. in Computer Science from BIHE University (Baha’i Institute for Higher Education), Iran in 2006 and M.Sc in Software Engineering from BIHE University, Iran in 2010. She is currently a researcher in the Center for Telecommunications and Multimedia at INESC TEC and pursuing her Ph.D in the MAP-i Doctoral Programme in the Faculty of Engineering of University of Porto. Her research interest include network management, data mining, probabilistic modeling and anomaly detection in IEEE 802.11 based wireless networks.

Ricardo Morla is an assistant professor of electrical and computer engineering at the University of Porto and principal investigator at INESC TEC, Portugal. His research interests are in the area of automatic system management with an emphasis on probabilistic modeling, prediction, anomaly detection, and root-cause analysis for ICT systems including network infrastructure and smart environments. He holds a Ph.D in computer science from Lancaster University UK.