Outlier Detection in 802.11 Wireless Access Points Using Hidden Markov Models

***Abstract*—In 802.11 Wireless Networks, detecting faulty equipments, poor radio conditions, and modification of user behavior through anomaly detection is of great importance in network management. The traffic load and users' movements on different access points (APs) in a wireless covered area vary from time to time, making these network management tasks harder. We intend to inspect the evolving structure of wireless networks and their inherent dynamics in order to provide models for anomaly detection. For this purpose we propose to explore the temporal usage behavior of the network applying various types of Hidden Markov Models. We observe the usage pattern of up to 100 APs in one week period in 2011 at the Faculty of Engineering of the University of Porto. The early stage of this study consists of constructing various Hidden Markov Models from the usage data of WLAN APs. We perform statistical techniques for outlier detection and justify the presented outliers inspecting the models' parameters and a set of HMM indicators. We finally introduce examples of wireless networks anomalous patterns by HMM states' transitions and provide an analysis of the entire set of APs under study.**

I. Introduction

The question of performance in 802.11 wireless networks becomes increasingly important as many new emerging applications such as mobile information access, real-time multimedia communications, and cooperative work require sufficient bandwidth and consistent connectivity. Some major circumstances leading to performance degradation in such networks are contention and collision, rate diversity and fairness, random losses and TCP performance and traffic asymmetry [9], [8]. Due to such fundamental issues of the wireless medium, users of 802.11 networks experience a number of connectivity problems such as authentication failures, intermittent connections to 802.11 APs and inconsistent or lack of coverage. These connectivity problems can be the consequence of RF interference, weak RF or RF holes, and users associating to overloaded APs [2].

To detect and address such anomalies in 802.11 networks we apply Hidden Markov Models which are frequently used in network measurements to obtain temporal information of signals in the network. Hidden states of HMMs encode different probability distributions. For example, states for *low* and *high* network activities and their respective observation distributions. HMMs also learn the probability of observing a hidden state in the next time slice given the current state. A statistical learning methodology estimates automatically the states' parameters and the state transition probabilities from the previous observations.

The typical approach for using HMMs in network-related work is to automatically learn an HMM for each behavior or class of network activities. The formed HMMs can be utilized in many different ways in different applications. For this work we detect anomalous time instances by analyzing the likelihood series of three types of HMM: a single model for all the network data, separate models for each AP's data, and groups of HMMs by mixture estimation technique. We justify and evaluate the anomalies detected by each model through HMM parameters exploration and analysis. Furthermore, we propose a number of network anomalous patterns deduced from states transition sequences and present the potential HMM variations capable of detecting specific types of patterns.

The rest of the paper is organized as follows. Section II presents related work. Section III describes the data set and the extracted features. In section IV HMM modeling and various HMM approaches are explained. Section V describes the outlier detection techniques and section VI provides a discussion on outliers analyzing the HMM parameters and introducing HMM indicators. Section VII deals with an example case and the experimental results of the entire set of data. Finally in section VIII conclusions and some future lines of work are presented.

II. Related Work

Hidden Markov Models have been applied to several network modeling and management studies for the purpose of estimation, classification, characterization and prediction of wired and wireless network features. Among the early lines of work is learning the probability of packet loss in Internet communication channels by observing the HMM states switches and inference about the state of the channel [14]. In [15], HMMs are utilized to model loss and temporal delays in network communication channels and a modified version of Expectation-Maximization (EM) algorithm is employed to learn parameters of the equivalent HMMs. Moreover, the significance of the hidden states is considered to be associated to particular congestion levels of the network, which is the similar intention of HMM states in our work. A typical well-known application of HMMs in network management is the traffic classification and identification by using packet-level statistical properties such as Packet Size (PS) and Inter Packet Time (IPT) [4], [5], [20].

The HMM dominance in learning temporal characteristics of various types of network makes it extremely interesting in anomaly detection work. In [17] for example, the observation sequences and state transitions are utilized to predict intrusion-prone state sequences where the HMM is built over the network traffic dataset. In another related research, network anomaly detection is performed by HMMs and Genetic Algorithm (GA) is involved in automating the training process of HMMs [7]. A recent related line of work [18] modeled time-varying people distribution by HMMs with outlier detection goal on people trajectories. In a similar manner to our work,

they studied the collective behavior of a large number of people to obtain information on unusual events and abnormal patterns. In Intrusion Detection research [19] HMMs are trained based on profiling system call sequences and shell command sequences. To discriminate the anomalous trends, likelihood of the sample sequences are computed and compared to a threshold. One major problem with this approach is the lack of generalization to support the users who are not uniquely identified by the system under consideration.

To the best of our knowledge, Hidden Markov Models have never been applied for addressing performance issues and connectivity problems of 802.11 wireless networks specifically on network management data like RADIUS [1]. In this work, we model the AP daily activities and recognize the abnormal hours in terms of connection loss and performance barriers. Three variations of HMM are provided and anomalous hours are marked exploring the HMM likelihood series. The justification of outliers is provided by HMM parameters analysis (HMM indicators). In addition to representing suspicious hours, a number of abnormal patterns are proposed as unlikely sequences of state transitions and a cross-check is performed between the anomalous hours and observed patterns. Finally, the best HMM variations are introduced for detection of specific kinds of outliers or abnormal patterns.

III. Data Set Description

The raw data set of this study consists of the daily summary of connections between 100 APs and their corresponding stations in the Eduroam hotspots at the Faculty of Engineering of the University of Porto (FEUP). The data used for the current work belongs to a one week period from 2011-12-12 to 2011-12-19. The entire data set incorporates records of 802.11 mobile stations' association to APs stored at a RADIUS authentication server. When a client associates/disassociates to an 802.11 AP, a *Start*/*Stop* event is recorded. A probing log event *Alive* is generated every 15 minutes while the client is still connected to the network [1].

A. Features

User Count: the number of unique users observed in a specific location (indicated by an AP) in a temporal period.

Sessions: the raw count of active sessions during a time-slot regardless of the owner. This attribute shows the number of attempts made to associate with an AP by the participating users.

Sightings: the number of probing event occurring every 5 minutes as an indicator of users' duration of stay in a location, and a counter for sessions longer than 5 minutes.

IV. Hidden Markov Models

A. Training HMMs

An n-state HMM is composed of:

- A set of states $S = \{s_i\}$, $1 \leq i \leq n$
- A state transition matrix $A = \{a_{i,j}\}$, $1 \leq i, j \leq n$
- The observation probability distributions (or emission matrix), $B = \{b_k\}$, $1 \leq k \leq m$
- An initial probability distribution $\Pi = \{\pi_i\}$, $1 \leq i \leq n$
- m = number of observation symbols in discrete models
- n = number of hidden states

Note that in continuous emissions, instead of having m outcomes for the observations, distribution parameters need to be determined. The set $\lambda = (A, B, \Pi)$ completely defines an HMM [13]. Using the model λ, an observation sequence $O = o_1, o_2, ..., o_T$ is generated as follows:

1) Select an initial state, s_i, according to the initial state probability distribution, Π;
2) Set $t = 1$;
3) Choose O_t according to observation probability distribution in state i, b_t;
4) Choose s_{i+1} according to the state transition probability distribution for state i, $a_{i,i+1}$
5) Set $t = t + 1$; return to step 3 and continue until $t > T$

In consonance with our data set, the models are multivariate, having 3 main features, and contain continuous Gaussian distribution. Hence the emission matrix B is defined by the distribution parameters, mean and covariance matrix associated with the set of states. In the proposed model, the HMMs contain fully connected states, hence transitions are allowed from any state to any other state.

B. Various Approaches to HMM Modeling

HMMs consist of states that encode different probability distributions. For the current work, we considered 3 states of *low*, *medium* and *high* addressing the usage performance of the APs and the respective observation distributions.

In this section three different HMM modeling approaches are provided to characterize the usage pattern of the entire set of APs employing various portion of the data. All types of HMMs in this work retain 3 states of *low*, *medium* and *high*.

1) Separate Models per AP: To induce HMMs specifically for each AP, a vector quantization process is performed on the set of observation sequences and 3 clusters produced (by k-means). Hence, the emission matrix of each model is formed by estimation of distribution parameters of each cluster (mean and covariance matrix). Moreover, the transition probability matrix is made by enumerating the sequences of observed states differentiated by the assumed parameters.

2) Single Model for all APs: In this approach it is assumed that the observation sequences of each AP can participate in generating a particular single HMM which characterizes the entire set of data in one model. The process of the formation and estimation of states' distribution parameters and transition probabilities are similar to the previous model with this difference that the states distributions and the transition probabilities are estimated using the expanded set of data. The single model is expected to be more robust using the data belonging to all APs, while the separate models might be more specifically generated for each AP.

3) Groups of APs and Mixture of HMMs: Provided a source of time series data, it is often advantageous to determin whether there are qualitatively different regimes in the data and characterizing those regimes. HMMs have been shown empirically to be capable of modeling the structure of the generative processes underlying numerous types of real-world time series. The mixture modeling is based on the well-established method of Expectation Maximization (EM) for estimating mixture parameters from the set of data. A mixture model [11] is defined as:

$$P(O_i|\Theta) = \Sigma_{k=1}^{K} \alpha_k P(O_i|\lambda_k) \quad (1)$$

The mixture probability density function (pdf) is parameterized by $(\alpha_1, ..., \alpha_k, \lambda_1, ..., \lambda_k)$, consist of the prior probabilities $\alpha_k, k = 1, ..., K$, and the likelihood function of the HMMs denoted by $P(O_i|\lambda_k)$. The λ_k is the set of parameters that describe the density functions of linear HMMs with multivariate emission distributions. The observed data O_i then corresponds to the multi-dimensional time-series that reflect the underlying usage pattern of the APs. The goal is to maximize Equation 1 by choosing optimal parameter set. This problem is generally solved by the EM algorithm which finds a local optimum for the above function. The outcome is the groups of APs with one optimized HMM as the representative of the group. The mixture method concisely performs the following main steps, given a collection of K initial HMMs $\lambda_1^0, ..., \lambda_K^0$:

1) **Iteration**:
   - Generate the initial groups of sequences by assigning each sequence O_i to the model k for which the likelihood is maximal.
   - Calculate new parameters for each model $\lambda_1^t, ..., \lambda_K^t$ using re-estimation algorithm (*Baum-Welch*) based on their current parameters $\lambda_1^{t-1}, ..., \lambda_K^{t-1}$ and the assigned weights of the participating sequences.
2) **Stop**: If the improvement of the objective function is below a given threshold ϵ, the grouping of the sequences does not change or a given iteration number is reached.

We presume that grouping APs and presenting an optimized HMM per group, has the potential to enhance the quality of the group models to conquer the weak points of the very generalized or very specific models (single HMM vs. separate HMMs).

V. Outlier Detection Methods

Outlier detection, also referred to as anomaly detection, event detection, or deviant discovery, is the process of distinguishing observations that lie outside the regular pattern of a distribution and do not comply with the well-defined expected behavior. Outliers can occur for a number of reasons, however the presence of an outlier in most cases indicates some sort of problem, where the outlier dose not fit the model under study or it is the consequence of an error in measurement.

Outlier detection techniques are generally categorized under unsupervised learning methodologies which deal with unlabeled data. In some circumstances, where acquiring labeled data is troublesome or rather time-consuming, unsupervised techniques are most widely applicable to obtain profound knowledge and discover underlying patterns of the data. In the current case, there is not any explicit ground truth provided for the RADIUS dataset which conveniently leads us to the anomalies. Therefore, we made use of some well-known outlier detection techniques as well as the various HMM approaches to investigate this demanding network management work.

A. Univariate Outliers: Feature by Feature

Individual data instances which are not compatible with the normal pattern of the rest of the data are called point anomalies. Point anomalies are the simplest form of anomalies detected as they lie outside the boundary of the normal zones. They can be single points each with a different pattern or small regions composed of several point anomalies.

To detect univariate outliers, features are inspected one by one without considering any correlation between them. Thus, any instance out of the normal boundary could be marked as an outlier without looking at the two other accompanying features. In this work, univariate outlier detection is performed using *boxplot.stats* function from R, which returns the statistics for producing *boxplots*. It labels the data points lying beyond the extremes of the *box-and-whisker* plot. An argument of coefficients is used to control how far the whiskers extend out from the box of a boxplot. We assumed $coef = 3$ to get the most extreme values as outliers. For the sample AP (AP#0) there is no outlier detected in the first two features, and only 1 extreme is observed in the third feature.

B. Multivariate Outliers: 3D Impression of Data

Assuming the data is multivariate normally distributed in d dimensions, the *Mahalanobis* distance of such set of data follows a *Chi-Square* distribution with d degrees of freedom. There are two approaches for outlier detection using Mahalonobis distances. The first one marks observations as outliers if they exceed a certain quantile of the chi-squared distribution. The second is an adaptive procedure looking for outliers specifically in the tails of the distribution, beginning at a certain *chisq-quantile*.

For this purpose the *aq.plot* function from R is used which plots the ordered squared robust Mahalanobis distances of the observations against the empirical distribution function of the MD_i^2. The distance calculations are based on the MCD estimator [6]. Figure 1 shows the outliers detected based on the assumed quantile (97.5%) and the adapted quantile for the 3D data of AP#0. The points labeled as *X* are the outliers projected on their two most robust principal components. The first and second approaches detect 2 and 3 outliers respectively. The multivariate outliers of the current work are identified using the first approach.

C. Temporal Outliers: Time Series

Contextual or conditional anomalies occur when a data instance is anomalous in one specific context and not in the others. The notion of context (or vicinity) determines the structure of data which has to be considered joint with data attributes to distinguish the anomalies. A salient example of

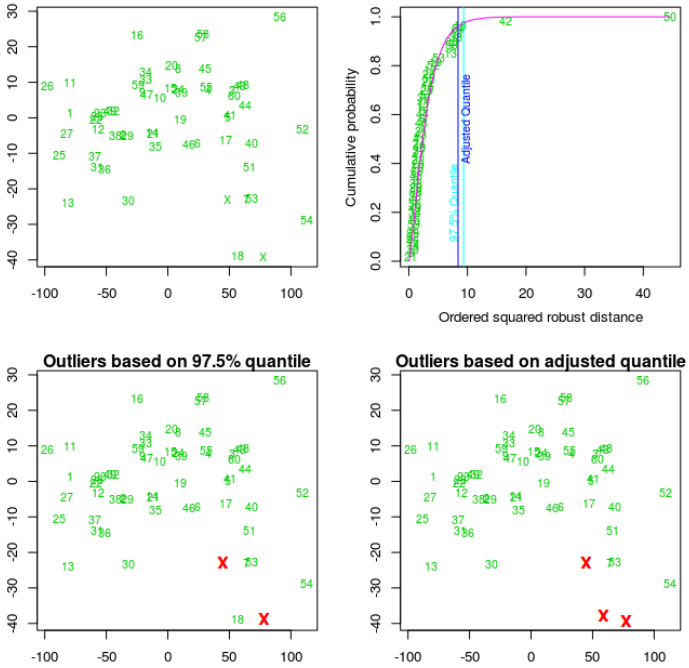


Fig. 1. Multivariate Outliers Detected for 3D Data of AP#0

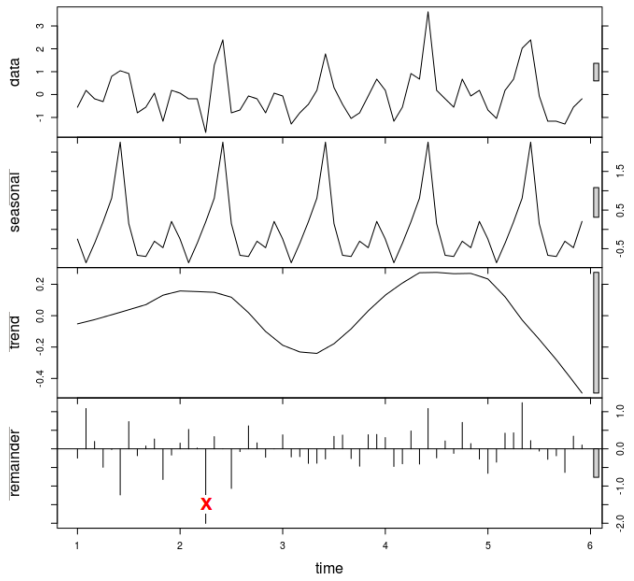


Fig. 2. Temporal Outliers Detected for Data of AP#0

a contextual attribute forming the data context is time in time series data. Contextual anomalies have been mostly explored in temporal data [10], [16] and spatial data [3].

Accordingly, the third type of outliers inspected for the current data set is the temporal outliers. Thereupon, *stl* function from R is practiced which decomposes a time series into seasonal, trend and irregular components using *loess* smoother. As the data set of this work is inherently periodic, containing 12 hours a day for 5 consecutive working days, a time series object with frequency of 12 is built and given to *stl* function. Figure 2 shows the time series data of AP#0, the seasonal, trend and remainder component. The *X* point is the only temporal outlier detected for the data of AP#0.

D. Hidden Markov Models: Likelihood Series

Let λ be an HMM and $O = o_1, o_2, ..., o_T$ a sequence of observations. Estimating the probability of observing O, assuming that it is generated by λ, $P(O|\lambda)$ is called likelihood. The likelihood values are expressed as logarithmic numbers to compensate the very small fractions caused by several probabilities' multiplications. Larger values of likelihoods (less negative) indicate the higher probability of an observation sequence generated by the given model. Measuring the likelihood of each observation instance (o_t, $1 \leq t \leq T$) given the HMM, produces a likelihood series of the entire observation sequence. In the likelihood series some values are basically out of the normal range of the rest of the series, hence simple outlier detectors indicate them as outliers. Figure 3 demonstrates the likelihood series of AP#0 for the separate, single and mixture HMM models.

VI. Discussion on Outliers

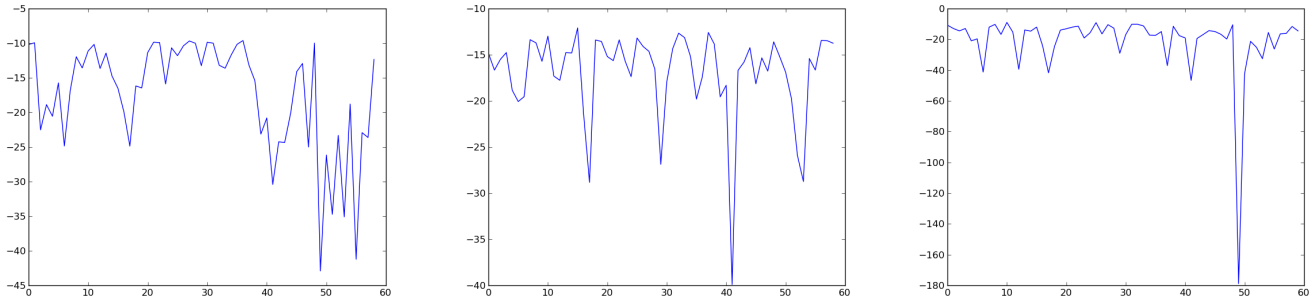
A. Outlier Quality Indicators

In this section some of the salient properties of HMMs are investigated which explain the likelihood results of different HMM models and provide some justifications to substantiate the detected outliers.

1) Large Distance to the adjacent HMM State: Given an HMM λ and an observation sequence of $O = o_1, o_2, ..., o_T$, the most probable set of states are generated by *Viterbi* algorithm as $O = q_1, q_2, ..., q_T$, $q_i \in S$. To estimate the distance of a data record in time t to its closest HMM state (q_t) in Viterbi path, we employed the concept of distance of a data point to a distribution. For this purpose *Mahalonobis* distance is calculated for each data point to its equivalent assigned state in the Viterbi path. To highlight the isolated records in terms of separation from the appointed HMM states, the univariate outlier detection method is utilized.

2) Less Likely State Transition: The highest transition probabilities are observed between identical states (s_i to s_i), while the lowest probabilities often occur between the most distant states (s_0 and s_2 in a 3 states HMM). The medium state (s_1) is the most uniformly distributed state in terms of transition probabilities to the higher (s_2) and lower (s_0) states. Having studied the regular HMM state transitions of the HMM variations, the least frequent transitions are more likely to be among the anomalous instances. For example when rare state modifications appear in Viterbi path or there exist no transition where it is expected to be according to the trained model. To distinguish the least likely state transitions, the HMM transition matrix is analyzed and transitions probabilities below 10% are marked.

On the other hand, the *less frequent state transitions* are also considered to be a sign of an unusual event. For example if in a Viterbi path the transition from state s_0 to state s_1 occur only once (out of 60 transitions), regardless of the probability



(a) Separate HMMs (b) Single HMM (c) Mixture of HMMs

Fig. 3. Likelihood Series of Three Variations of HMMs

of the transition, it is very rare and thus an outlier-prone transition. This indicator could be different from the *less likely transitions* as it comes from the actual Viterbi sequences, not just the model parameters.

3) HMM States' Separation: Continuous HMM states, indicate various distributions represented by distribution parameters. To conceive the distances between HMM states, we utilized the *bhattacharyya.dist* function from R calculating the distance between distribution pairs which provides the potential of testing the severity of the observed anomalies. For example, successive fluctuation between states could be a symptom of network mal-functioning. However, if the distance between the two alternating states is not large enough the possibility of a network connectivity problem is less and it might be because of the users mobility to and from an AP.

Other advantage of studying distance between HMM states is to understand the separation model of states and declare the anomalous patterns with more confidence. For example when in a model (s_1) and (s_2) are very close, the anomalous pattern of *2020* is almost identical to the *1010* pattern. So not observing the former one does not mean that there is no such anomalies, and the similar pattern should also be taken into consideration.

It must be carefully considered that HMM indicators and HMM outlier detectors deal with two different aspects of the HMM models. HMM outliers come from the abnormal likelihood values in the likelihood series, which reflect the trait of each and every observation instance as well as the overall quality of the model. For example if the likelihood value of a unique observation is out of the normal boundary it could be due to the observation issues or the inability of the model to provide a reasonable likelihood for such emission values. In any case, the extreme observations are marked as outliers because of their likelihood values. The quality of the models is evaluated observing the outcomes of various HMM models and make a inter-model comparison. On the other hand, HMM indicators monitor the characteristic of each data instance concerning the model parameters which determine whether there is any sign of abnormality associated with data points and assert the potential source of the problem from the model parameters' viewpoint. For example a data point is marked by *less likely state transition* HMM indicator because of its low transition probability which is one of the model parameters. Hence, according to the proposed model this point is suspicious. If the model has already report the same point as outlier (based on its likelihood value), then we claim a true match. The likelihood outliers that can not be confirmed by the model indicators are claimed to be false positives of the model. Note that although attributes and functions of the HMMs are compatible, they do not refer to the same prospect.

B. Anomalous Patterns: Collective Outliers

The normal usage pattern observed in APs consists of one or two peaks a day. Regardless of the AP location type (classroom, auditorium, administrative office, cafeteria, etc.) a gradual transition between hidden states is expected during the normal periods. There exist a number of anomalous patterns when a sequence of actions occur together. In HMM terminology, the anomalous patterns could be defined as the occurrences of a number of state transitions in a specific order. For example a high state transformation to a low state and remaining there for some successive time slices, could indicate a special type of anomaly interpreted by network experts. From the AP point of view, the observed anomalous patterns may fall into one of the following categories. [12]

1) AP Halt/Crash: when the HMM is in the low state after a drop from a higher state and remains low for some successive hours (*210000* or *110000* or *220000*). Such patterns indicate possibility of AP halt or crash.

2) Persistent Interferences: where repeated downturns are observed for a given AP in a day, possibly due to RF interference and RF holes. Decline from a higher state to the lowest state (s_0) more than three times a day indicates such anomalies (*10..10..10* or *20..20..20*).

3) AP Overload: in the presence of high HMM state right after a low state and intermittent fluctuation between the two states (*0202* or *0101*) in case states 0 and 1 are distant or *1212* in case state 1 and 2 are distant). This could be the case of heavy utilization by some few users.

4) Interferences across the AP vicinity: when the unlikely state modifications occur to neighboring APs within specified time intervals, it is possibly due to the increase in the number of collisions occurring in the wireless medium and interferences from the adjacent APs. To probe this type of anomaly in a precise manner, smaller granularities should be applied which is considered among the future work.

TABLE I. DETECTED OUTLIERS OF AP#0 BY ALL THE OUTLIER DETECTION TECHNIQUES

	Day 2												Day 4												Day 5											
Day Hours	8	9	10	11	12	13	14	15	16	17	18	19	8	9	10	11	12	13	14	15	16	17	18	19	8	9	10	11	12	13	14	15	16	17	18	19
Univariate																		X																		
Multivariate																		X								X										
Temporal				X																																
Separate HMMs																																				
Dist Ind.	*																									*						*				
Prob Ind.																																				
Trans Ind.																																				
Single HMM																		X																		
Dist Ind.	*																																			
Prob Ind.																																				
Trans Ind.																																				
Mixture of HMMs																		X								X										
Dist Ind.						*												*													*					
Prob Ind.																										*										
Trans Ind.																										*	*									

TABLE II. HMM OUTLIERS COMPLIANCES WITH STOA OUTLIERS AND HMM INDICATORS

	STOA Outliers			HMM Indicators		
	Univariate	Multivariate	Temporal	Large Distant (Dist)	Low Probability (Prob)	Rare Transition (Trans)
Separate Model	29%	70%	74%	15%	9%	10%
Single Model	29%	60%	73%	22%	8%	24%
Mixture Model	21%	81%	76%	24%	12%	15%

VII. EXPERIMENTAL RESULTS

In this section two principal lines of the current study are discussed: HMM outliers and the anomalous patterns. The experimental results are presented through an example study, and a systematic analysis of the entire set of data. The summary of the relevant tests are provided in Table I-IV

A. HMM Outliers

1) An Example: In this part one AP is selected (AP#0) and its outliers detected by various detection techniques (state of the art (STOA) and HMM models) are analyzed and evaluated. Table I provides a comparative outlook of all detected outliers of the example by different models. The learning period consist of 5 working days and 12 hours per day, from 8 a.m to 7 p.m. No outlier was observed in day 1 and day 3, therefore the 3 other days are selected to be represented in this table. The HMM indicators are also exposed right below each of the HMM models.

There are some hours which are marked by a majority of the detectors and there are instances that are pointed out only by one or two detectors. The 5^{th} data point on day 4, at 13:00 is labeled as outlier by 4 of the 6 detectors. Inspecting the data, the third feature (user count) is out of the normal range, so the entire record is marked by univariate detector. It is also detected by multivariate technique which consider the Chi-Square of Mahalonobis distances, but not by the temporal detector. The mixture HMM detects this point as well as the single HMM, while the separate model do not.

The mixture model report this data instance of day 4 as an outlier due to its large distance to the assigned hidden state in Viterbi path, but there is no reasoning found by the single model indicators explaining why this point is detected as an outlier. As the table results show, none of the single indicators support this decision. The mentioned data point is very likely to be an outlier as the majority of detectors acknowledge that, but the single model, in this case, is not competent enough to justify its detection, and the separate model do not recognize it at all. Several number of fluctuations, as Figure 1 (a) shows, disable the separate model to distinguish the extreme likelihoods, hence this model provides no outlier for this example. It is only the mixture model which detects the point and can support its decision. There is another outlier proposed by the mixture model at the 2^{nd} time slice of the day5 at 9:00 which is supported by two of the HMM indicators, the low probability and the rare transition between hidden states. Among the STOA detectors, the multivariate technique also marks this point as outlier. None of the single or the separate HMM variations are able to detect this point. Both data points detected by mixture model are compatible with the STOA outliers (one with univariate and both with multivariate), and they are both supported by the HMM indicators of the mixture model.

If the outliers can not be justified by the HMM indicators they are referred to as *soft false positives*. If they can not be explained neither by the HMM indicators, nor by the STOA outliers, they are declared as *hard false positives*. The unique outlier introduced by the temporal method, is not detected by any of the HMM variations. The overall compatibility of the HMM outliers with the STOA outliers and HMM supported indicators are summarized in Table II and III in the next section.

However, the quality of HMMs in modeling observation sequences is an important concern. How accurately an HMM, represents the characteristics of its associated observations and conquer the over-fitting issues at the same time? Basically, each HMM model produces a number of false positives which in the absence of the precise ground truth is complicated to distinguish. In spite of this, a background knowledge of the network abnormalities in addition to the HMM parameters inspection more in depth, provide reliable explanations to estimate anomalies and differentiate the true and false positives.

2) The Systematic Approach: In this section we present the likelihood results of three HMM variations for all the APs, comparing the set of outliers to STOA outliers and the HMM parameters indicators. Those HMM outliers not in agreement

TABLE III. COMPARISON OF HMM VARIATIONS

	STOA	HMM Ind.	Soft FP	Hard FP
Separate Model	81%	45%	55%	12%
Single Model	77%	57%	43%	15%
Mixture Model	85%	56%	44%	6%

with any of the HMM indicators, marked as *Soft FP*. If they are not compatible neither with HMM indicators nor with the STOA outliers they are referred to as *Hard FP* in Table III that signifies there is no reasoning understood behind their selection as outliers.

This experiment also reveals the best HMM models for different types of anomalies. For instance Table II shows that mixture HMM is the best model to capture multivariate outliers, while in detecting other types of the STOA outliers all three types of HMMs perform almost identical. The mixture model is more capable of recognizing outliers caused by low transition probabilities and large distances to the proper hidden state, while the single model is the most efficient model for outliers due to rare state transitions distinguished in Viterbi path.

Table III shows that more than 77% of the outliers detected by three versions of HMM are in accordance with the STOA outliers, while the second column displays that around half of the HMM outliers can be validated by HMM indicators (large distance, low probability and rare transition). The higher the compliance between HMM outliers and HMM indicators, the more confidence is assured in the presented anomalies to the network administrating team. The single and mixture models demonstrate a superior rate of compliance to HMM indicators, while the ratio of the *Hard FP* in the mixture model still remains lower than the two other HMM models. It must be noticed that STOA outliers are not utilized as ground truth, but just as auxiliary verification tools providing an extra level of certainty for explaining the HMM outliers.

B. Anomalous Patterns

Table IV presents the observed anomalous patterns in the separate, single and mixture HMM models by the percentage of the patterns' occurred per AP and per day, in addition to the patterns' compliances to the observed outliers.

Among the three categories of anomalous patterns, the least frequent one observed is *Persistence Interference* which is the repetition of downturns to the lowest state (*10* or *20*) more than three times a day. Generally the state transition of *20* is very rare and its three recurrences per day, is only captured once by the mixture model (in 1 hour of 1 day). However, transition from the medium state to the low state (*10*) is more frequently identified specifically by the separate model.

In the case of *AP Halt/Crash*, *110000* pattern appears more than the two other patterns (*210000* and *220000*). However the most salient patterns of this category is *220000*, which demonstrate a sudden decline of usage from a peak that lasts for 4 consecutive hours. It should be noted that such types of anomalies needs a further control by the network administrator. For example if this pattern happens in classroom or auditorium, it can be due to the termination of a crowded event which several number of participants decide to disconnect their wireless stations and leave the place almost concurrently.

Another interesting outcome is inspecting the pattern observations by various HMM models. For example when the single HMM detects (*220000*) pattern, it is extremely acceptable to be a true catch, while detecting (*110000*) pattern by the same model could contain many false alarms. The reason behind that is the high probability of watching only 2 states (*0* and *1* or *1* and *2*) in Viterbi path created by the single model, due to the HMM state generalization (expanded HMM states generated from the entire dataset). The high rate of observing *110000* pattern by the single model (in 40% of APs and in 13% of days) and low ratio of *220000* pattern observation (in 2% of APs and in 0.6% of days) affirms the above discussion.

The *AP Overload* pattern which is identified by the HMM state fluctuations, is very unlikely between the high and low states (*0202*) and is captured just once appeared in the separate model. The (*1212*) pattern seems to occur less than *0101* pattern, and the former one is a more appropriate estimation of *AP Overload*, as the extremely jammed hours (high HMM state) normally follow by medium states in AP overloaded situations and the intermittent fluctuation between these two states can be a typical form of this class of anomalies.

Table IV performs a cross-check between the distinguished anomalous patterns and the detected outliers of each HMM model. The *Outlier Compliance* row of this table shows in what portion of the observed anomalous patterns, an outlier is also detected by the model. For example in the mixture model, the second pattern of the *AP Overload* (*0101*) is observed in 10% of the APs, which in 4 of them (40%), an HMM outlier is also occurred during the pattern at the specified anomaly point. For each anomalous pattern there is an anomaly point (HMM state), which the AP is crashed or the interference is occurred or the AP is overloaded at that specific state. For instance in the *AP Halt/Crash* pattern, the third state is where the anomaly happens (downturn to the low HMM state). Observing the previous and next hours just determines the type of anomaly and affirms the required duration. The compliance between the anomalous patterns and the observed outliers of each model provides a higher level of confidence for the proposed anomalies to the network management team.

The comprehensive analysis of such anomaly-related patterns or even more complicated alternatives, informs the network managers about various types of anomalies, the level of severity, and the extent of confidence to the presented patterns. Such information permits the network managers to take immediate actions or make long-term plans for the maintenance or re-structure decisions for the network.

VIII. CONCLUSION AND FUTURE WORKS

The increasing significance of performance and connectivity issues of 802.11 wireless networks has made anomaly detection tasks a very crucial aspect of network management and administration. Analyzing the user's behavioral patterns and learning efficient models to detect the network deficiencies and anomalous patterns is the main contribution of this work. We proposed a new application of HMMs in performance anomaly detection of 802.11 wireless networks and presented several indicators of outliers by HMM parameters' analysis. Furthermore, we provided a number of anomalous patterns associated with such networks in terms of HMM state transitions.

TABLE IV. ANOMALOUS PATTERNS OBSERVED IN 3 HMM VARIATIONS

		AP Halt/Crash			Persistence Interference		AP Overload		
		210000	110000	220000	10 (3x)	20 (3x)	0202	0101	1212
Separate Model	AP (%)	10	25	12	14	0	1	18	8
	Day (%)	2.2	6.4	2.4	3	0	0.2	4	2.4
	Outlier Compliance (%)	10	20	33.3	64.2	0	0	5.5	12.5
Single Model	AP (%)	2	40	2	3	0	0	4	1
	Day (%)	0.4	13	0.6	0.6	0	0	0.8	0.2
	Outlier Compliance (%)	100	52.5	100	0	0	0	25	0
Mixture Model	AP (%)	4	31	7	9	1	0	10	1
	Day (%)	0.8	8	1.6	2.2	0.2	0	2.4	0.2
	Outlier Compliance (%)	25	22.5	28.5	33.3	0	0	40	0

The experimental results show that HMM models are able to discover a portion of the STOA outliers (univariate, multivariate and temporal), while introducing some additional outliers which can be justified by HMM parameters indicators (large distance to assigned HMM state, low transition probability, and rare state modification). The single and mixture models outperformed the separate HMMs in terms of accuracy and HMM indicators conformity.

In future work we intend to decrease the granularities from one hour to 10-15 minutes to provide more complicated and more accurate anomalous patterns while keeping the complexity of HMM models low. A very promising line of work is to improve the HMM models, specifically the mixture estimation, in order to improve the potential of such models in detection of more authentic outliers with less false alarms. Meanwhile, we work on arranging some network measuring and assessment approaches for RADIUS data which hopefully will provide us with a notion of ground truth and will make the evaluation process more straightforward.

REFERENCES

[1] Rfc 2139 radius authentication, 2013.

[2] A. Adya, P. Bahl, R. Chandra, and L. Qiu. Architecture and techniques for diagnosing faults in ieee 802.11 infrastructure networks. In *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, MobiCom '04, pages 30–44, 2004.

[3] Fabrizio Angiulli and Fabio Fassetti. Detecting distance-based outliers in streams of data. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, pages 811–820. ACM, 2007.

[4] A. Dainotti, W. de Donato, A. Pescape, and P. Salvo Rossi. Classification of network traffic via packet-level hidden markov models. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–5, 2008.

[5] A. Dainotti, A. Pescape, P. Salvo Rossi, G. Iannello, F. Palmieri, and Giorgio Ventre. Qrp07-2: An hmm approach to internet traffic modeling. In *Global Telecommunications Conference, 2006.*, pages 1–6, 2006.

[6] P. Filzmoser and M. Gschwandtner. *mvoutlier: Multivariate outlier detection based on robust methods*, 2013. R package version 2.0.3.

[7] J.J. Flores, A. Antolino, and J.M. Garcia. Evolving hmms for network anomaly detection - learning through evolutionary computation. In *Networking and Services (ICNS)*, pages 271–276, 2010.

[8] A. Gupta, J. Min, and I. Rhee. Wifox: Scaling wifi performance for large audience environments. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 217–228, 2012.

[9] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11b. In *INFOCOM 2003. IEEE Societies*, volume 2, pages 836–843 vol.2, 2003.

[10] H. Jiawei. Outlier detection for temporal data: A survey. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 25(1):1, 2013.

[11] G. MacLachlan and D. Peel. *Finite mixture models*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, USA., 2000.

[12] D. Massa and R. Morla. Abrupt ending of 802.11 ap connections. In *IEEE Symposium on Computers and Communications*, 2013.

[13] L. Rabiner and B.-H. Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

[14] K. Salamatian and S. Vaton. Hidden markov modeling for network communication channels. *SIGMETRICS Perform. Eval. Rev.*, 29(1):92–101, 2001.

[15] P. Salvo Rossi, G. Romano, F. Palmieri, and G. Iannello. Joint end-to-end loss-delay hidden markov model for periodic udp traffic over the internet. *Signal Processing, IEEE Transactions on*, 54(2):530–541, 2006.

[16] R. Andrew Weekley, Robert K. Goodrich, and Larry B. Cornman. An algorithm for classification and outlier detection of time-series data. *Journal of Atmospheric and Oceanic Technology*, 27(1):94–107, 2010.

[17] C. Yang, F. Deng, and H. Yang. An unsupervised anomaly detection approach using subtractive clustering and hidden markov model. In *Communications and Networking in China, 2007. CHINACOM '07*, pages 313–316, 2007.

[18] S. Yang and W. Liu. Anomaly detection on collective moving patterns: A hidden markov model based solution. In *Internet of Things (iThings CPSCom), 4th International Conference on Cyber, Physical and Social Computing*, pages 291–296, 2011.

[19] DY. Yeung and Y. Ding. Host-based intrusion detection using dynamic and static behavioral models. *Pattern Recognition*, 36(1):229–243, 2003.

[20] C. Yin, S. Li, and Q. Li. Network traffic classification via hmm under the guidance of syntactic structure. *Computer Networks*, 56(6):1814–1825, 2012.