Anomaly-based Intrusion Detection Using Mobility Profiles of Public Transportation Users

Jeyanthi Hall, Michel Barbeau, Evangelos Kranakis
Carleton University, School of Computer Science
Ottawa, Ontario, Canada K1S 5B6
Telephone: 1-613-520-4333
Email: jeyanthihall@rogers.com, {barbeau,kranakis}@scs.carleton.ca

***Abstract*— For the purpose of anomaly-based intrusion detection in mobile networks, the utilization of profiles based on hardware signatures, calling patterns, service usage and mobility patterns have been explored by various research teams and commercial systems, namely the Fraud Management System by Hewlett-Packard and Compaq. This paper examines the feasibility of using profiles, which are based on the mobility patterns of mobile users, who make use of public transportation, e.g. bus. More specifically, a novel framework, which makes use of an instance based learning technique, for classification purposes, is presented. In addition, an empirical analysis is conducted in order to assess the impact of two key parameters, namely the sequence length and precision level, on the false alarm and detection rates. Moreover, a strategy for enhancing the characterization of users is also proposed. Based on simulation results, it is feasible to use mobility profiles to enhance anomaly-based intrusion detection in mobile wireless networks, provided that the user mobility profiles adequately reflect the mobility behavior of users.**

Keywords: Mobile Networking, Security, Intrusion Detection, IBL, and Mobility Profiles. [1]

I. Introduction

Mobile wireless networks continue to be plagued by theft of identity and intrusion. Both problems can be addressed in two different ways, either by misuse detection or anomaly-based detection. Misuse detection is carried out by recognizing instances of well known patterns of attacks. The main limitation of this approach is that the system fails to uncover new kinds of attacks unless it has been instructed to do so. Anomaly-based intrusion detection (ABID) consists of observing and recognizing deviations from normal behaviour, which has been captured and maintained in electronic profiles. It is generally acknowledged that the main limitation of the anomaly-based detection approach is that it generates a higher rate of false positives than the misuse detection approach.

The limitation imposed by anomaly-based detection approach can be minimized by combining observations across time and across domains. When intrusion detection is carried out using a given profile, multiple observations can be correlated in time using a state-probabilistic model such as Bayes filters [1]. This strategy accommodates a moderate degree of variability in normal behaviour, as indicated by Morin and Debar in [2], and consequently reduces the rate of false alarms. Furthermore, using a statistical tool such as multivariate analysis [3], the detection results, associated with multiple profiles from different domains, can also be combined to further reduce the rate of false alarms. Examples of intrusion detection systems (IDSs), which make use of multi-sensor data for enhanced detection, include AAFID by Balasubramaniyan [4] and EMERALD by Porras and Neumann [5].

The use of different profiles for ABID has been investigated by various groups. Node/device profiles are created by exploiting the unique hardware signature of their wireless interface [6], operating system (proposed by Taleck [7]) and other characteristics of a wireless device. In terms of user-based profiling, the use of calling patterns for fraud detection in cellular networks is explored by Boukerche et al. [8]. Calls are classified into the normal category or anomalous category based to whether or not the time and location of the calls match the profile of the user. If the probability of fraud is high, then a warning message is sent to the client who owns the phone.

Commercial systems, namely the Fraud Management System by Hewlett-Packard (FMS-HP) [9] and Compaq (FMS-C) [10] also make use of service usage profiles, which are built using calling patterns, call frequency, call times and duration, wireless home/roaming behaviour and other call-related information. Although both FMSs offer some services, which permit them to be differentiated, they both detect multiple types of fraud by examining all calls (e.g. streams of call detail records used for billing purposes) and other-related events (event records).

Indeed, an intrusion unfolds in many aspects of a network. Referring to the ISO/OSI seven-layer model, anomaly detection in a communication system can be conducted from the application down to the physical layer. Research in the area of network routing misbehaviour detection has been conducted, for example, by Just, Kranakis and Wan [11] and Zhang and Lee [12]. At the link layer, medium access control misbehaviour detection, has been investigated by Kyasanur and Vaidya [13]. Work associated with identity theft detection, at the physical layer, has been carried out by Hall, Barbeau and Kranakis [14].

[1]The authors graciously acknowledge the financial support received from the following organizations: Alcatel, Mathematics of Information Technology and Complex Systems (MITACS) and Natural Sciences and Engineering Research Council of Canada (NSERC).

In this paper, we examine the feasibility of using profiles, which are based on the mobility patterns of users, for ABID at the application layer. In particular, a novel framework that makes use of a statistical classifier is presented. The instance based learning (IBL) classification system [15] used is a general class of machine learning techniques. In addition, we focus on the analysis of two key system parameters, namely the sequence length and precision level, in order to determine their impact on the false alarm and detection rates. The mobility behaviour of users is also taken into consideration. A strategy for enhancing the characterization of users is also proposed. Finally, simulations, which were conducted, are based on location broadcasts (LBs) from users, who make use of public transportation, e.g. bus, in the area of Los Angeles. The high density of these users promotes a high probability of intrusions, a necessary prerequisite for a meaningful analysis.

Our objective is to supplement existing user and device-based profiles, with those based on mobility, in order to further enhance ABID in mobile wireless networks. Moreover, the use of mobility profiles is particularly applicable for addressing the problem of stolen cell phones, given that the mobility behavior of the thief and the authorized user are likely to be different. Lastly, we believe that the underlying framework can be applied, with minimal translation (e.g. use of cells instead of geographical coordinates), to the mobile wireless network.

The remaining sections of the paper are organized as follows. Section 2 presents the framework for the application of mobility profiles to intrusion detection. Whereas Section 3 discusses the analysis of the two key system parameters, simulation results are presented in Section 4. Other related work are identified in Section 5, followed by the conclusions and future research initiatives in Section 6.

II. ABID USING MOBILITY PROFILES

This section provides an overview of the ABID system, which makes use of mobility profiles of authorized users. As with most IDSs, the two key objectives are to define the user mobility profiles (UMPs) and to design an appropriate classification system.

A. Framework

Details of the framework, which is used for the implementation of the ABID system, are provided in this subsection. It is important to note that the detection process, as described in the sequel, is applied to each authorized user. Moreover, during the profiling phase, the subset of the activities, from data collection to the definition of the UMP, is typically carried out on a one-time basis and prior to classification. However, in order to address the issue of concept drift, where the mobility patterns of users change with time, it is essential that the profiles be updated periodically. One approach is to maintain a window of the training patterns that is continuously shifted in time as new sequences are added (analogous to the use of exponentially weighted moving average). As the window is shifted, the definition of UMP is updated accordingly. This should reduce the rate of false alarms and correspondingly increase the detection rate.

The intrusion detection process begins with the data collection exercise. Once the LBs, which contain location coordinates (LCs) and other data, have been captured for a period of 3-6 months, a high-level mapping (HLM) is applied. The objective of the HLM is to decrease the granularity of the data in order to accommodate minor deviations or intra-user variability between successive LBs. Specifically, a mapping from a LC with high granularity to a cluster-based (lower granularity) model is used. Upon completion of this phase, the LCs (feature) are extracted from each broadcast during feature extraction. A set (defined by sequence length) of these chronologically-ordered LCs are subsequently concatenated to define a mobility sequence. This process continues until all the mobility sequences (data set) have been created. The unique sequences (training patterns) from the first four of the six partitions of the data set is stored in the UMP, along with other user-related information. During the classification phase, an observed set of mobility sequences of a user is compared to the training patterns in his/her profile. If the average similarity measure to profile (SMP) value falls within the pre-established thresholds, the mobility sequences are considered normal, otherwise a flag is raised.

B. High-Level Mapping

The term *intra-user* variability refers to the difference in the LCs (j represents the latitude and i represents the longitude) that are transmitted by user A as he/she travels using routes one (solid line) and two (dashed line), see Fig. 1. So, for example, if a LC, in the area of $(j+5, i+4)$, is sent while on routes one and two, these coordinates could potentially be different. Let us assume that the full sequence of LCs, associated with route one, have been captured and stored in the profile (training patterns) of user A. If the sequence of LCs, which are transmitted while user A is on route two, is compared to those in the training patterns, it would result in a similarity value of zero. Thus, altering the representation of the sensor data (e.g. LCs), using an appropriate HLM scheme, becomes necessary in order to decrease the granularity of the LCs. In another words, the HLM converts LCs, which are in close proximity, to a cluster (size of cluster is based on one of three precision levels). As a result, the similarity between the corresponding LCS in two sequences is increased.

This mapping process, which is applied to the LC in each LB, is carried out as follows. The original format of the LC is (###.#####) and (###.#####), where the first and second terms (###.#####) represent the latitude and longitude respectively. Based on the precision level (PL), the LC is truncated and rounded to the specified number of digits after the decimal point. For example, with level three, the specified digit of the first and second terms (###.##) is rounded to 0 if it is within 0-4 and to 5 if it is within 5-9 range. Thus, for example, the LC 33.14623,114.26874 is mapped to 33.10,114.25. Similarly, the HLM for levels two and one are (###.#) and (###.0)

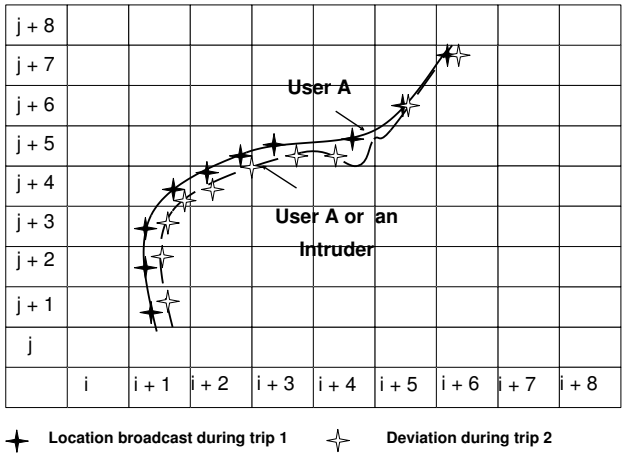

Fig. 1. Intra-user and inter-user variability

respectively. The choice of PL or cluster size is explored in Section 3, see Fig. 1.

Caution must nevertheless be exercised since minimizing intra-user variability will also minimize *inter-user* variability (deviations between location broadcasts from different users). As depicted in Fig. 1, the same logic would apply to potential intruders as well, resulting in a potentially successful impersonation attempt.

In general, inter-user variability must be maximized in order to correctly distinguish between legitimate users and intruders.

C. Feature Extraction

The extraction of LCs (feature) from the HLM data is required in order to create mobility sequences. A mobility sequence is defined as a sequence of LCs. The selection of the appropriate sequence length is also addressed in Section 3.

The feature extraction process concatenates the first set (e.g. ten) of chronologically-based LCs into a single sequence. Furthermore, each subsequent sequence is created by using the LCs at the $i+1$ to $j+1$ indexes of the HLM data stream, where i (LC1) and j (LC10) represent the first and last LC of the first sequence. Hence, a sequence of ten location coordinates is obtained by shifting i and j by one, as suggested by Lane and Brodlay [16]. The purpose of using an overlapping window (shifted by one) is to accommodate different sequences that begin with different LCs. In other words, it permits each location coordinate to become a starting point of a sequence.

This process is repeated until all the LCs in the data stream have been exhausted. The resulting set of sequences, henceforth referred to as original sequences, serves as input to the profile definition and classification phases.

The use of LCs only in the current feature set is intentional and supports our objective of determining the maximum success rate possible. Additional features, such as timeframe, will be investigated in the future for maximizing the *inter-user* variability.

D. Profile Definition

Once the mobility sequences have been obtained, the next step is to create the UMPs. A detailed description of each component in the UMP ensues.

Identifier represents the unique identification of the user, which has been issued by Industry Canada. It is transmitted with all LBs. *Training Patterns* characterize the mobility behaviour of a user. Due to factors, such as traffic and weather, a mobility sequence of a user may deviate from the norm. This deviation is referred to as noise, which must be minimized. The term *window size* refers to the number of mobility sequences to be used for obtaining the average or noise-suppressed NSMP value. In addition to the reduction of noise, the size of the window also influences the length of time, corresponding to the number of location coordinates, required before a detection verdict can be rendered. While the window size is identical for all users (in this iteration), it is feasible that this value can be customized to reflect the level of noise within the mobility patterns of a given user. One possible strategy for identifying the level of noise or intra-user variability is to determine the number and frequency of unique location sequences in the training data. Whereas a small number of unique sequences with high frequencies supports the notion of consistent behaviour, the reverse exemplifies a more chaotic behaviour. Whether or not these mobility sequences reflect normal behaviour is based on the *minimum* and *maximum* thresholds. If the NSMP value falls within the thresholds, it is considered normal, otherwise, a potential intrusion is suspected. The values of the thresholds are determined by obtaining a distribution of the NSMP values, between the training patterns and parameter sequences (5^{th} partition of the data set), and by applying the desired false alarm rate (application-dependent) to the distribution.

E. Classification

The final step, in anomaly-based detection, is the classification of an observed behaviour as normal or anomalous.

As stated earlier, during the classification phase, an observed set of mobility sequences of a user is compared to his/her profile, which contains a set of training patterns. For each mobility sequence being compared to the training patterns, the maximum similarity value (discussed in the sequel) is obtained. If the average of these values falls within the pre-established thresholds, then the user is considered legitimate, otherwise a flag is raised. The following subsection provides a brief overview of the key concepts defined in IBL. Readers are encouraged to consult the paper by Lane and Brodlay [16] for a more detailed discussion of the IBL framework.

The IBL framework requires the application of key concepts, which are enumerated in Table I.

Similarity Measure

As you may recall, a mobility sequence is composed of a chronologically ordered sequence of LCs and that these sequences are used for training, establishment of parameters and test/simulation (final partition of the data set) purposes.

TABLE I

KEY CONCEPTS ASSOCIATED WITH IBL CLASSIFICATION

Concepts	Description
Similarity Measure (SM)	Similarity between a test and a training pattern
Similarity Measure to Profile (SMP)	SM between a test and all training patterns
Noise Suppression	Mean SMP for 10 consecutive test sequences
Decision Rule	Classification of test sequence as normal or anomalous

Therefore, the *similarity* of two sequences X (from the set of test sequences) and Y (from the set of training patterns) of equal length l is defined as follows:

$$sim(X,Y) = \sum_{i=0}^{l-1} w(X,Y,i)$$

with:

$$w(X,Y,i) = \begin{cases} 0 & \text{if } i < 0 \text{ or } x_i \neq y_i \\ 1 + w(X,Y,i-1) & \text{if } x_i = y_i \end{cases}$$

where i represents the index of the sequence of LCs. Thus $w(X,Y,i)$ equals zero if the LCs of the X and Y sequences at index i are not identical. Otherwise, a value of one is added to the outcome of $w(X,Y,i)$ at $i-1$.

Similarity Measure to Profile

As aforementioned, a user profile D contains user-related information, which includes a set of training patterns. Whereas the SM is determined based on a one to one comparison of the LCs of a test sequence and training pattern, the SMP is calculated by performing a one to many comparison of an observed test sequence X with *all* the training patterns in a profile D. It is defined as:

$$sim_D(X) = max_{Y \in D} sim(Y,X).$$

The maximum value of $sim_D(X)$ is:

$$\sum_{i=1}^{l} i = \frac{l(l+1)}{2}.$$

Thus, the SMP is the maximum of the SM values.

Noise Suppression

In the subsection on feature extraction, the notion of using each LC (from a long stream of coordinates) as a starting point i of a sequence of length l was introduced. This form of segmentation results in a set of sequences (original sequences), whereby a sequence, starting at location coordinate i, is called the i-th sequence.

As with all chaotic systems, noise is inherent and reflects the deviation of a test sequence from the patterns stored in the profile. A degree of *intra-user* variability is to be expected, since it is a function of many factors including traffic conditions and weather. Nevertheless, noise can be suppressed, to some extent, by calculating the average SMP of a set of W test sequences, where W represents the size of the window. Thus, the average SMP over a window of length W ending at position i is defined as:

$$v_D(i) = \frac{1}{W} \sum_{j=i-W+1}^{i} sim_D(j).$$

The term $v_D(i)$ is referred to as the *noise-suppressed SMP (NSMP) value*.

Decision Rule

Whether or not a given set of test sequences exhibit normal mobility behaviour can be determined by comparing the resulting NSMP value to the pre-established minimum t_{min} and maximum t_{max} thresholds. While t_{min} is used to detect sequences, which have low NSMP values, t_{max} proves beneficial in detecting sequences that have unusually high similarity to the profiled behaviour, perhaps an indication of a replay attack.

The calculation of t_{min} and t_{max}, for each user, is carried out by applying an acceptable false alarm rate r (application-specific) to a normalized probability distribution (NPD) of NSMP values. Thus, t_{min} and t_{max} are dependent on r and NPD.

The parameter r dictates the width of the acceptance region (between t_{min} and t_{max}) on the NSMP axis, see **??**. It represents a trade-off between false alarm and detection error rates. Hence, a smaller value of r corresponds to a wider acceptance range. As a result, the rate of false alarms is decreased. However, the increased acceptance region also causes the detection error rate to increase.

As far as the NPD is concerned, it is obtained by using the parameter sequences and the training patterns, obtaining a distribution/histogram of NSMP (in the range of $0, \ldots, l(l+1)/2$) and normalizing this distribution based on the probability of each NSMP value.

Finally, t_{max} and t_{min} are established using $r/2$ quantiles (upper and lower) of the NPD, as proposed by Lane and Brodlay [16]. The number of sequences and the actual sequences (training vs parameter) used for the calculation of NPD are important factors to be considered. As far as the number of sequences are concerned, it is dependent on the variability of the original LCs and the level of HLM used to minimize this variability. Using a high level of HLM (coarse granularity) results in location coordinates being more similar, and thus, reduces the number of sequences in the training set. The number of sequences to be used in the parameter set is not as significant so long as they reflect the mobility behaviour of the user.

As to which sequences, from the initial set of sequences, should be used for training, parameter and test data represents a more challenging problem. One option, which has been implemented in this iteration, is to divide the initial set of sequences into partitions of $4/1/1$ with the first $4/6$ of the

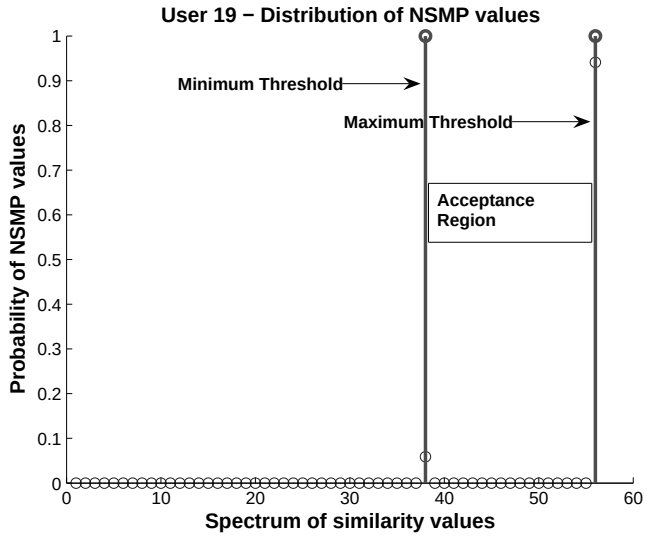

Fig. 2. Minimum/Maximum thresholds

sequences (with respect to time) being allocated to the training data, followed by $1/6$ to parameter and the last $1/6$ to test data. By allocating the first and the largest set of sequences to training, the probability of accurately characterizing the mobility behaviour of a user is increased. This is, of course, based on the assumption that the mobility patterns of a user is typically established within a given timeframe. The shape of the NPD reflects the accuracy with which the mobility behaviour of a user has been characterized. Regardless of the strategy being used, it is clear that the profile of users, which includes the set of training patterns, must be updated on a continual basis. Moreover, a replacement strategy, which favors the most recent patterns, should be employed in order to limit the storage space and to maintain an acceptable level of performance (currently being investigated).

Fig. 2 illustrates the application of $r = 0.05$ to the NPD of user 19, who was selected at random. In this figure, the x-axis represents the spectrum $(0, \ldots, l(l+1)/2)$ of the similarity values that are possible for a sequence of LCs of length 10. Please note that the actual values are in the range of $(1, \ldots, l(l+1)/2) + 1$ for improved graphical representation. The y-axis represents the probability of each NSMP value in the normalized probability distribution. Both the minimum and the maximum thresholds are indicated using vertical lines. What is illustrated in the figure is the width of the acceptance region (from the minimum threshold to the maximum threshold), which is a function of the NPD and the false alarm rate r. The narrow acceptance region, located at the higher end of the spectrum, is a desirable property. In particular, the location of the minimum and maximum thresholds at NSMP values of 38 and 56 respectively, reflects the accuracy of mobility characterization. As a result, the true detect rate should increase while the detection error rate should decrease correspondingly.

III. Empirical Analysis of System Parameters

In the previous sections on HLM and feature extraction, we had indicated that the two key parameters, namely the cluster size or PL and sequence length, are of significance and that an appropriate value had to be selected. The PL, used in HLM, determines the degree to which the intra-user variability is being minimized in order to reduce the number of false alarms. On the other hand, sequence length not only specifies the number of LCs in a mobility sequence, but more importantly, the maximum similarity value attainable for a given length.

Aside from stating the obvious, our first objective is to determine the impact of these parameters on the characterization of users (distribution of the NSMP values) and detection errors or intrusions (successful impersonation attempts against a user). We address the impact of these parameters on false alarm and detection rates in the section on simulation.

Given that the mobility behavior of the 50 users does differ to some extent, and that this variability is likely to influence the analysis of both parameters, we have categorized these users based on the precision with which the training patterns are being followed (repetitions). The three classes are defined as follows. Whereas class one represents users with the highest level of similarity (consistent behaviour), class two and three are associated with those with progressively lower levels of similarity (chaotic behaviour). Due to space constraints, we focus on the results obtained for user 19 (class 1 with 40% of users) as they illustrate the expected behaviour, associated with proper characterization. Nevertheless, we briefly comment on results (figures not shown) obtained for user 23 (class 2 with 56%) and user 41 (class 3 with 4%).

A. Sequence Length

Fig. 3 illustrates the use of three different lengths (5,10,15) for sequences and the impact on the characterization of user 19. Values of NSMP, which are located at the lower-end of the SM spectrum are vulnerable to the choice of r. Since r dictates the width of the acceptance region, in particular the minimum threshold, all values of NSMP that are less than the threshold are treated as false alarms.

Other static parameters used include the window size of 100, cluster size with the precision level of one (only the integer portion of the LCs were used), and minimum threshold of two. The maximum threshold, however, was based on the sequence length being used.

In Fig. 3, the x-axis represents the range or spectrum of similarity values for a given sequence length. However, since the results, associated with each length, have been incorporated into one plot, the range of the x-axis is actually from 1-121. In other words, results obtained for length of five are localized towards the lower end of the similarity value spectrum. NSMP values, which have been normalized, are indicated by the y-axis.

What is being illustrated is as follows: as the sequence length is increased, the percentage of NSMP values, located at the higher-end of the SM spectrum starts to decrease. In this case, the NSMP values are located precisely at 15, 55 and 120

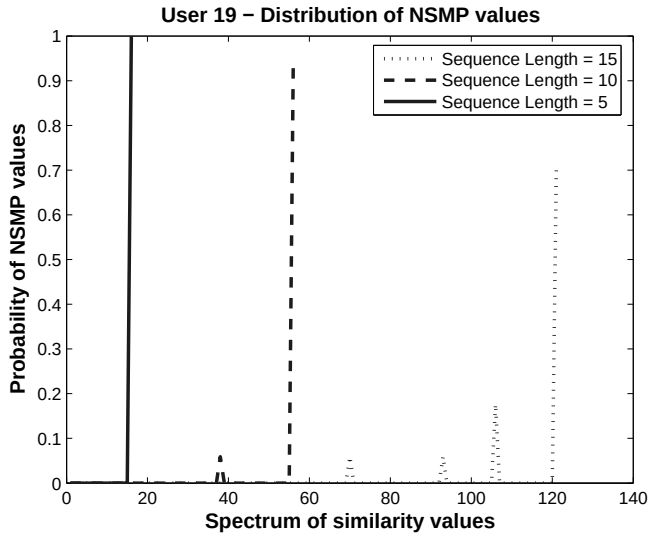

Fig. 3. Characterization of mobility behaviour

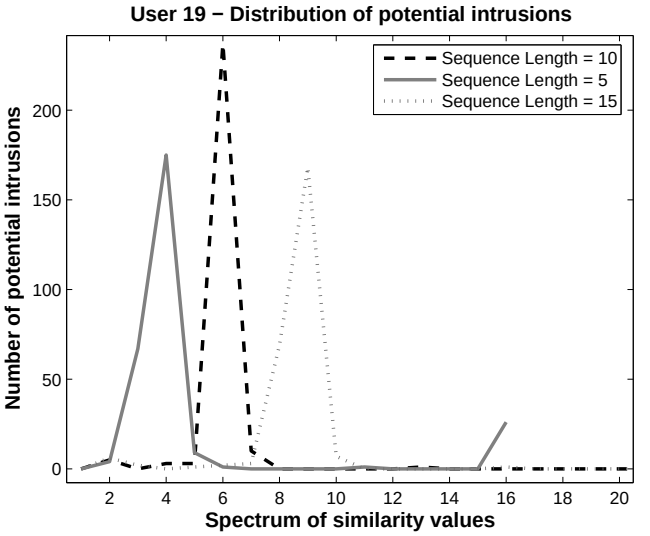

Fig. 4. Potential intrusions

on the x-axis. Furthermore, as the percentage of these values decreases, they are distributed towards the lower end of the spectrum. This behaviour is logical since the probability of achieving a high NSMP value decreases as the sequence length is increased. Therefore, should the NPD of a user be localized at the higher end of the spectrum, selecting a larger sequence length would be not be advisable since it shifts the NPD further towards the left. However, if the NPD is located at the lower end of the spectrum (user 41), it is advantageous to use a larger sequence length, since this results in the NPD being shifted towards the higher end of the spectrum. On the other hand, when the NPD is distributed between the lowest and highest similarity values (user 23), a larger sequence length is also desirable for shifting the NPD towards the center of the spectrum.

We continue our analysis of the impact of sequence length on the distribution of intrusions (detection errors). All parameters, which were used in the previous test, remain the same, with the exception that the NSMP values of potential intrusions, are calculated using the training data of user 19 and test data from the remaining 49 users.

Fig. 4 depicts the distribution of intrusions associated with each of the three sequence lengths used. It is important to note that we have emphasized the range of SM values between 1-16, since most of the intrusions are located in this range. The original x-axis does cover the range of 1-121. This figure demonstrates the fact that, as the sequence length is increased, the distribution shifts towards the higher end of the SM spectrum. This behaviour is justified since there is a higher probability of achieving a high NSMP value when the sequence length is longer. In terms of users 23 and 41, the key difference is the magnitude of the distribution. Due to the more chaotic behaviour, the magnitude is higher for user 23 and even more so for user 41.

The last detail to note is the number of intrusions at location 16 on the x-axis. It is an indication that one or more of the 49 users have mobility patterns that are identical (based on the HLM being used) to user 19. In fact, most of these intrusions are caused by user 13. Varying the cluster size to increase the granularity of the LCs, discussed next, addresses this problem.

B. Precision Level

We proceed with the analysis of the PL and its impact on the characterization of users and number of potential intrusions. Given that our focus is to minimize the number of intrusions first and then address the problem of characterization, we have used a sequence length of 5.

Fig. 5 indicates that the distribution of NSMP, associated with a given precision, shifts towards the lower end of the spectrum as the PL or granularity is increased, e.g. from PL2 to PL3. This behavior is consistent with all three classes of users. Therefore, a lower PL (larger cluster size) can be used for HLM in order to improve characterization. Using a lower PL decreases the distance between similar LCs. Thus, the probability of a match between a training pattern and a parameter sequence is higher, resulting in higher NSMP value.

Although the use of a smaller PL is desirable for characterization purposes, it becomes problematic where intrusions are concerned, see Fig. 6. What is evident, in this figure and applicable to all classes of users, is that the distribution shifts towards the lower end of the spectrum as the PL is increased (smaller cluster size). Moreover, the intrusions at SM value of 15 are eliminated. This should not come as a surprise since increasing the PL also increases the distance between two LCs. As a result, the probability of obtaining a high NSMP value is reduced, as indicated by the distribution of intrusions for PL3. Thus the use of a smaller PL would result in an increase in the detection error (intrusion) rate and a corresponding decrease in the detection rate.

In summary, the selection of values for both the sequence length and cluster size is a challenging task since all of the

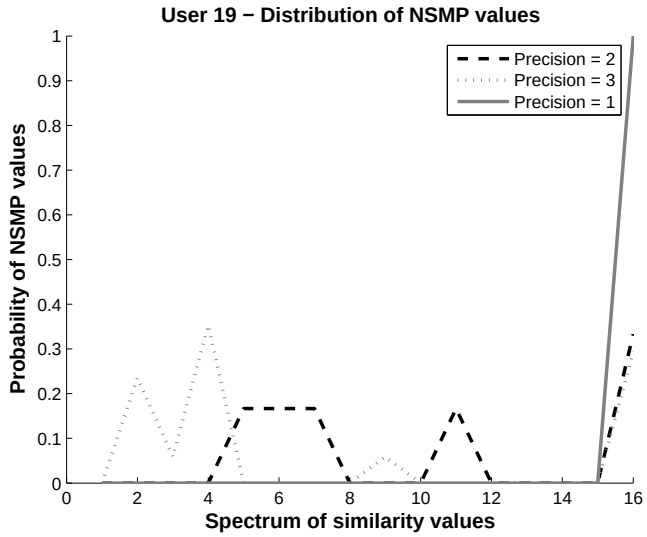

Fig. 5. Precision level and characterization

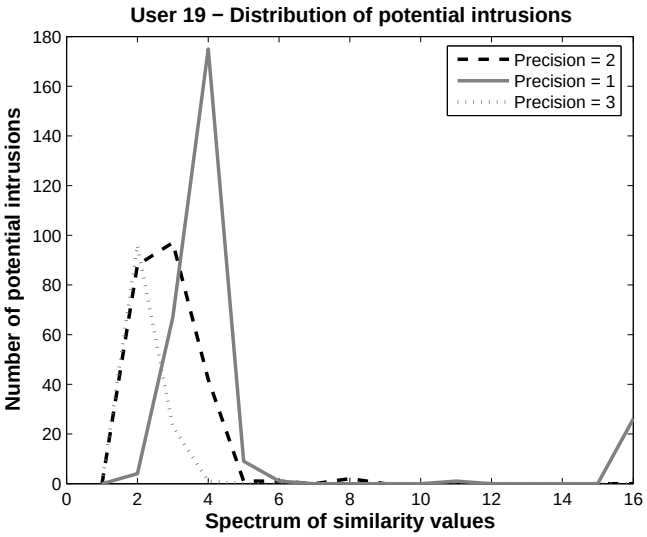

Fig. 6. Precision level and intrusions

possible values produce results that are negatively correlated.

IV. Simulation

The primary objective of the simulation exercise was to determine the impact of PL on the false alarm and detection rates (metrics). We relaxed the use of various sequence lengths for the time being, given that a smaller sequence length is preferable for for improving the detection rate. We were also interested in the correlation between the quality of characterization, attainable using IBL, and the resulting false alarm and detection rates.

A. Simulation Infrastructure

Details of the simulation infrastructure are as follows. The acquisition of the LBs was carried out using the Automatic Position Reporting System (APRS) and appropriate hardware (e.g. receiver and antenna). The APRS is an internet-based system (open-source) that tracks objects and users using amateur radio.

It has been specified by Markoulidakis in [17] (follow-up on the UMTS RACE specification) that nearly 50% of all mobile users use public transportation, e.g. bus, and that they can be characterized. Furthermore, this statistic has been confirmed to some extent by Wu in [18]. Hence, we targeted users who took the bus in the area of Los Angeles. This city was selected due to the high density of APRS users. Finally, the top 50 users (those who had transmitted the highest number of LBs) were selected to participate in the simulation.

The captured LBs (approx. 2 million) were transferred from the APRS to a MySQL database for further processing. All subsequent analysis and simulation were carried out using Matlab software.

B. Details of Simulation

The simulation exercise was carried out for each of the 50 profiled users in the IDS. In order to determine the percentage of false alarms, a comparison or classification was made between the sequences in the test data of user A and his/her training patterns. The resulting NSMP values, which were outside the minimum and maximum thresholds (r=0.05), were considered false alarms (FAs). As with the FAs, the percentage of true detect (TD) (detection) was obtained by comparing the test sequences of the remaining 49 users to the training patterns of user A. The resulting NSMP values, which fell outside the thresholds (r=0.05), were considered TDs. Statistics, corresponding to the metrics, were obtained for all profiled users.

C. Simulation Results

We limit the discussion and focus on the results obtained for the representatives of each class, namely users 19, 23 and 41. Although an attempt was made to generalize the results for each class, the end result did not fully highlight the nuances found within each class.

False Alarm and Detection Rates

Fig. 7 illustrates the percentage of FAs and TDs corresponding to the three PLs.

We begin with the discussion of user 19 (class 1) and observe that there are no FAs for all three PLs. This is due to the fact that the three minimum thresholds of (16,5,2) associated with PLs 1,2, and 3, see Fig. 5, are all greater than the value of one. This is an indication that the mobility sequences in the test data are similar to those in the parameter data, which had been used to establish the thresholds. In terms of TDs, the percentage of TDs decreases as the PL is increased. Further scrutiny reveals that this behaviour is appropriate in light of the fact that the distribution of NSMP values shifts to the lower end of the SM spectrum, see Fig. 5. Therefore, as the minimum thresholds shift towards the lower end of the SM spectrum, the probability of intrusions, within the acceptance

range, is higher. This results in an increase in the detection error rate and a corresponding decrease in the TD rate.

The characterization of user 23 (class 2), on the other hand, is not as optimal. In fact, the NSMP values are distributed between the SM values of 1 and 16 (figure not shown) for the PL of one. The wide acceptance region and the fact that the minimum threshold has a value of one (actual value is zero) reflects the absence of sequences (parameter data) in the training data. Although the test sequences may or may not be similar to those in the parameter data, all of them have fallen within the thresholds, resulting in zero FAs. These two factors (wide region and value of minimum threshold) have also permitted all intrusions to take place resulting in a TD rate of zero. As the precision level is increased to PL2 and the maximum threshold becomes equivalent to the minimum threshold, it becomes more evident that the test sequences are dissimilar to those in the parameter data, but are nevertheless similar, to some degree, to those in the training patterns. As a result, the FA rate becomes 100%. The corresponding TD rate at PL2 also increases due to the fact that the intrusions, which fell outside the minimum and maximum threshold of one, are now being detected at this level. Finally, as the PL is increased to three, the number of FAs decreases as a result of the increase in intra-user variability between the test sequences and the training patterns. As expected, the TD rate also decreases as the PL is increased. Simply stated, the increase in the inter-user variability has caused fewer intrusions (being considered detections) to take place.

Results for user 41 (class 3) are very interesting, although somewhat misleading. We observe that, as with user 19, there are zero FAs for all three PLs. However, unlike user 19, the minimum and maximum thresholds of one and four respectively, for all precision levels, has permitted all test sequences to fall within the narrow acceptance region. Similarly, the minimum threshold of one has also permitted all intrusions to take place, even when the test sequences of all other users are dissimilar to the training patterns of user 41.

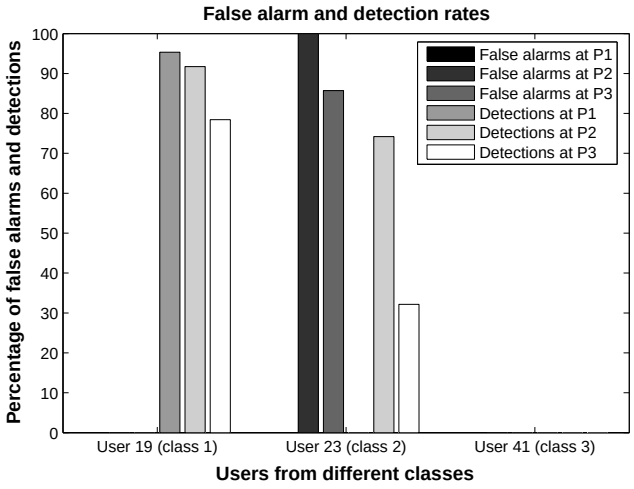

Fig. 7. False alarms and detections for different precision levels

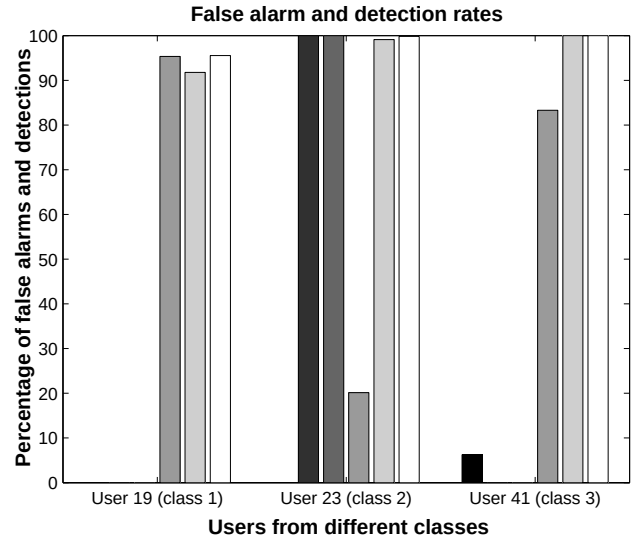

Fig. 8. False alarms and detections using enhanced characterization

Enhanced Characterization

What is clearly evident, from the previous simulation exercise, is the need to shift the minimum threshold towards the higher end of the spectrum, such that it is greater than one. One simple strategy is to add the mobility sequences in the parameter data, which have a similarity value of one, to those in the training data. This strategy reduces the width of the acceptance region and shifts the NPD, especially the minimum threshold, towards the higher-end of the spectrum.

Fig. 8 demonstrates the application of this strategy and the resulting impact on the FA and TD rates. With user 19 (class 1), the FAs remain the same whereas the TD rates (for all PLs) are increased, as expected. Moreover, the largest increase of 19% is associated with PL3, a desirable outcome. As far as user 23 (class 2) is concerned, the three TD rates, associated with PL1,PL2 and PL3 are increased by 20%, 33% and 233% respectively. However, the FAs for PL3 is also increased due to the dissimilarity of some of the test sequences to those in the parameter set. Finally, the results for user 41 (class 3) exemplify the effectiveness of this strategy. Although a 5% increase in the FAs (at PL1) has been incurred, there is, nevertheless, a significant improvement in the TDs (85%, 100%, 100%), associated with the three PLs.

V. Related Work

The use of user mobility profiles for ABID in mobile networks has not been researched extensively. However, research initiatives, which have been undertaken by researches include Buschkes, Kesdogan, and Reichl [19], Samfat and Molva [20] and Sun and Yu [21]. The work conducted by Buschkes makes use of sequences of cells traversed by users as a feature of the profile. Intrusion detection of users, using cloned phones, is carried out by analyzing major deviations from the route. Similarly, the behaviour of users is modeled based on the telephony activity and migration patterns by Samfat and

Molva. The implementation of multi-level intrusion detection, at the visitor location and using multiple profiles, differentiates their work from the others. Finally, the most recent work by Sun and Yu also makes use of sequences of cells as a feature. However, the characterization is accomplished via a high order Markov model [22]. Furthermore, the sequences, which are stored in a mobility trie (an acceptable solution given that the size of the alphabet is small) is updated using the technique of Exponentially Weighted Moving Average.

Of course, user mobility profiles have also been used to address the inefficiencies of location-area based update schemes. Details can be found in the work by Wong [23] and Ma [24]. Finally, the use of profile-based protocols for enhanced routing in wireless Mobile Ad Hoc Networks is addressed by Wu in [18].

VI. Conclusions and Future Research Initiatives

Based on simulation results, it is feasible to use mobility profiles for enhancing ABID in mobile wireless networks, so long as the mobility behaviour of users has been accurately characterized. One simply strategy, which enhances the characterization of users and increases the detection rate at a minimal cost (low percentage of FAs), is to incorporate the missing sequences, from the parameter set into the training set. Furthermore, the issue of concept drift (accommodating variability in mobility behavior over time) can also be addressed by continuously monitoring the false alarm rate and selectively incorporating newly observed mobility sequences into the training set, using a window that is shifted in time (analogous to exponentially weighted moving average). The selection criteria can be based on pre-established thresholds, such as the frequency of sequences encountered.

Once the characterization of users has been adequately addressed, the selection of specific values for sequence length and precision level should be based on the level of intra-user variability, which can be specified in the user's profile. Categorizing users into different classes, based on the level of variability, represents an alternate strategy.

Finally, the use of the IBL classification technique, within the framework, is suitable since the definition of the similarity measure is comparable to that of the euclidian distance. Supplemented by the high level mapping exercise, which reduces the intra-user variability between mobility sequences and training patterns, this technique performs well, as indicated by the false alarm and detection rates obtained for all three classes of users.

As far as future research initiatives are concerned, the following issues will be explored in the near future: user privacy; concept drift; the expansion of the feature set, to include timeframe and other relevant features, for improving detection rate; a comprehensive analysis of the system performance for comparison purposes; and the allocation of different parameter values to users based on their mobility behaviour.

Acknowledgment

The authors wish to thank Andrew Robison and Frederic Gariador, from Alcatel Canada, for fruitful discussions.

References

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall PTR, 2002.

[2] B. Morin and H. Debar, "Correlation of intrusion symptoms: an application of chronicles," in *Proceedings of the International Conference on Recent Advances in Intrusion Detection*, Berlin Heidelberg, 2003, pp. 94–112.

[3] J. Joseph, F. Hair, E. Anderson, W. Black, and R. Tatham, *Multivariate Data Analysis*. Prentice Hall PTR, 1998.

[4] J. Balasubramaniyan, J. Garcia-Fernandez, D. Isacoff, E. Spafford, and D. Zamboni, "An architechture for intrusion detection using autonomous agents," COAST Laboratory Purdue University, Tech. Rep., 1998.

[5] P. Porras and P. Neumann, "EMERALD: Event monitoring enabling responses to anomalous live disturbances," in *Proceedings of the Twentieth National Information Systems Security Conference*, 1997, pp. 353–365.

[6] M. Riezenman, "Cellular security: better, but foes still lurk," *IEEE Spectrum*, pp. 39–42, June 2000.

[7] G. Taleck, "Ambiguity resolution via passive os fingerprinting," in *Proceedings of the International Conference on Recent Advances in Intrusion Detection*, Springer-Verlag Heidelberg, 2003, pp. 192–206.

[8] A. Boukerche, *Security and fraud detection in mobile and wireless networks*. John Wiley and Sons, Inc., 2002, ch. 27.

[9] (2003) Hp - fraud management system. Hewlett Packard. [Online]. Available: http://www.hp.com

[10] (2001) Compaq - fraud management system. Compaq. [Online]. Available: http://www.hp.com/hps/nsp/

[11] M. Just, E. Kranakis, and T. Wan, "Resisting malicious packet dropping in wireless ad-hoc networks using distributed probing," in *Proceedings of 2nd Annual Conference on Adhoc Networks and Wireless (ADHOC-NOW'03)*, Montreal, Canada, 2003, pp. 151–163.

[12] Y. Zhang and W. Lee, "Intrusion detection in wireless ad-hoc networks," *Mobile Computing and Networking*, pp. 275–283, 2000.

[13] P. Kyasanur and N. Vaidya, "Detection and handling of mac layer misbehaviour in wireless networks," Digital Equipment Corporation, Tech. Rep., 2002.

[14] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *Proceedings of the 3rd IASTED International Conference on Communications, Internet and Information Technology (CIIT)*, St. Thomas, U.S. Virgin Islands, November 2004, pp. 201–206.

[15] D. Aha, D. Kibler, and M. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, 1991.

[16] T. Lane and C. Brodlay, "Temporal sequence learning and data reduction for anomaly detection," *ACM Transactions on Information and System Security*, vol. 2, pp. 295–331, August 1999.

[17] J. Markoulidakis, G. Lyberopoulos, D. Tsirkas, and E. Sykas, "Evaluation of location area planning scenarios in future mobile telecommunication systems," *Wireless Networks*, vol. 1, 1995.

[18] K. Wu, J. Harms, and E. Elmallah, "Profile-based protocols in wireless mobile ad hoc networks," *Local Computer Networks*, pp. 568–575, 2001.

[19] R. Buschkes, D. Kesdogan, and P. Reichl, "How to increase security in mobile networks by anomaly detection," in *Proceedings of the Computer Security Applications Conference*, Phoenix, AZ, USA, Dec. 1998, pp. 3–12.

[20] D. Samfat and R. Molva, "IDAMN: an intrusion detection architecture for mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1373–1380, Sept. 1997.

[21] B. Sun and F. Yu, "Mobility-based anomaly detection in cellular mobile networks," in *International Conference on WiSe 04*, Philadelphia, Pennsylvania, USA, 2004, pp. 61–69.

[22] L. Rabiner and B. Juang, *An introduction to hidden markov models*. Prentice Hall PTR, 1986.

[23] V. Wong and V. Leung, "Location management for next generation personal communications networks," *IEEE Network*, pp. 18–24, Sept. 2000.

[24] W. Ma and Y. Fang, "A new location management strategy based on user mobility pattern for wireless networks," in *Proceedings of the 27th Annual Conference on Local Computer Networks*, 2002.