Using Mobility Profiles for Anomaly-based Intrusion Detection in Mobile Networks

Jeyanthi Hall, Michel Barbeau, Evangelos Kranakis
Carleton University, School of Computer Science
Ottawa, Ontario, Canada K1S 5B6

I. Extended Abstract

The high rate of false alarms, which results from the use of anomaly-based intrusion detection (ABID) in mobile networks, can be addressed by combining observations across time and across domains. When ABID is carried out using a single profile, multiple observations can be correlated in time using a state-probabilistic model such as Bayes filters [1]. Furthermore, using a statistical tool such as multivariate analysis [2], the detection results, obtained using 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 et al. [3] and EMERALD by Porras and Neumann [4].

To date, 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, operating system (proposed by Taleck [5]) 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. [6]. In addition, commercial systems, namely the Fraud Management System by Hewlett-Packard (FMS-HP) [7] and Compaq (FMS-C) [8] also make use of service usage profiles.

The focus of this research is to examine the feasibility of using mobility profiles for enhancing ABID in mobile networks. In particular, a unique classification approach, using an instance based learning (IBL) technique [9], is adopted. In addition, we focus on the analysis of two key system parameters in order to determine their impact on the false alarm and detection rates. Finally, simulations, which were conducted, are based on location broadcasts (LBs) from users, who make use of public transportation, e.g. bus in Los Angeles. This environment promotes a high probability of intrusions, a necessary prerequisite for a meaningful analysis.

A. Intrusion detection using mobility profiles

As with most IDSs, the two key objectives are to define user mobility profiles (UMPs) and to design an appropriate classification system.

The intrusion detection process, which is repeated for each user, 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 location broadcasts. 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. Details of user profiling and classification

The mobility profiles are defined using the following parameters: identifier, training patterns, window size, and minimum and maximum thresholds. The *identifier* represents the unique identification of the user. *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. Whether or not these mobility sequences reflect normal behaviour is based on the *minimum* and *maximum* thresholds. 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.

As stated earlier, the classification process is carried out using the IBL technique. In brief, for each sequence being compared to the training patterns, the maximum similarity measure is obtained. A similarity measure between two sequences is calculated by not only comparing each corresponding element (e.g. LC) in the two sequences, but also taking into consideration the chronological sequencing as well. Finally, the NSMP value is obtained by calculating the average of the maximum similarity measures for a set of sequences.

C. Empirical Analysis of System Parameters

The two key system parameters, which influence the false alarm and detection rates, are the cluster size and sequence length. The size of the cluster, used in HLM, dictates the level of abstraction of the LCs and consequently influences the degree to which intra-user variability is minimized. On the other hand, sequence length not only specifies the number of LCs in a mobility sequence but more importantly, the maximum similarity measure for a given length.

A detailed analysis of both parameters indicates the following: As the sequence length and the cluster size are increased (independently), the rate of false alarms is decreased. However, the detection rate is also decreased, since there is a higher probability of two sequences (from different users) being similar.

D. Simulation

The key objective of the simulation exercise was to determine the correlation between the quality of characterization (the degree to which the mobility behavior of users is reflected in the training patterns) and cluster size. We relaxed the use of various sequence lengths for the time being.

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. 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.

In terms of the simulation exercise, it was carried out for each of the 20 profiled users in the IDS. Whereas the false alarm rate for each user was determined using his/her training patterns and test sequences (final partition of the data set), the detection rate was obtained by using his/her training patterns and the test sequences from all the remaining users.

In brief, the use of a smaller cluster size is preferable for increasing inter-user variability (improved detection rate). However, the prerequisite is that the mobility behaviour of users has been accurately characterized. Our effort to minimize the false alarm rate has resulted in 85% of the 20 users having a false alarm rate of 10% or less. However, only 45% of the users had a detection rate of 80% or more due to the inadequacy in characterization. One option for increasing the detection rate without further increasing the false alarms is to add the mobility sequences, from the parameter set, that are missing in the training set.

E. Related Work

The work conducted by Buschkes [10] 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 [11]. 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 [12] also makes use of sequences of cells as a feature. However, the characterization is accomplished via a high order Markov model [13].

F. Conclusion

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. Otherwise, the selection of specific values for key parameters, such as sequence length and cluster size becomes less meaningful.

One strategy (currently being investigated), for enhancing the characterization of users and addressing the problem of concept drift (keeping UMP up to date), is to maintain a window of the newly observed sequences (analogous to the exponential weighted moving average). These sequences can then be used to update the training patterns periodically. This should reduce the rate of false alarms and correspondingly increase the detection rate.

ACKNOWLEDGMENT

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). They also wish to thank Andrew Robison and Frederic Gariador for fruitful discussions.

REFERENCES

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

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

[3] 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.

[4] 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.

[5] 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.

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

[7] (2003) Hp - fraud management system. Hewlett Packard. [Online]. Available: http://www.hp.com

[8] (2001) Compaq - fraud management system. Compaq. [Online]. Available: http://www.hp.com/hps/nsp/

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

[10] 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.

[11] 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.

[12] 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.

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