JOURNAL OF KOREA MULTIMEDIA SOCIETY, VOL. 10, NO. 12, DECEMBER 2007(pp. 1726-1732)

Design and Evaluation of a Rough Set Based Anomaly Detection Scheme Considering the Age of User Profiles

Ihn-Han Bae†

ABSTRACT

The rapid proliferation of wireless networks and mobile computing applications has changed the landscape of network security. Anomaly detection is a pattern recognition task whose goal is to report the occurrence of abnormal or unknown behavior in a given system being monitored. This paper presents an efficient rough set based anomaly detection method that can effectively identify a group of especially harmful internal attackers - masqueraders in cellular mobile networks. Our scheme uses the trace data of wireless application layer by a user as feature value. Based on this, the used pattern of a mobile's user can be captured by rough sets, and the abnormal behavior of the mobile can be also detected effectively by applying a roughness membership function with the age of the user profile. The performance of the proposed scheme is evaluated by using a simulation. Simulation results demonstrate that the anomalies are well detected by the proposed scheme that considers the age of user profiles.

Keywords: Anomaly Detection, Rough Set, User Profile, Age

1. INTRODUCTION

The nature of mobile computing environment makes it very vulnerable to an adversary's malicious attacks. The use of wireless links renders the network susceptible to attacks ranging from passive eavesdropping to active interfering. Unlike wired networks where an adversary must gain physical access to the network wires or pass through several lines of defenses at firewalls and gateways, attacks on a wireless network can come from all directions and target at any node. Damages can include leaking secret information, message contamination, and node impersonation [1].

In general, two complementary approaches exist to protect a system: prevention and detection. Prevention-based techniques, like authentication and encryption, can effectively reduce attacks by keeping illegitimate users from entering the system. They are usually based on some symmetric and asymmetric mechanisms to assure that users conform to predefined security polices. Nevertheless, in cellular wireless networks, the mobile devices are not physically secure: they can be lost or stolen. Since tamper-resistant hardware and software are still too costly for most users, such insecurity makes all secrets of the device open to malicious attackers. An attacker, once possesses the device as well as all secrets associated with the device, he becomes an internal user and is able to cause great damage to the whole network. At this time, intrusion detection approaches, utilizing different techniques to model the users' normal behavior and system vulnerabilities, come into place to help identify malicious activities [2].

In this paper, we propose the intrusion detection technique that the normal profile of each mobile is constructed by using rough sets and anomaly ac-

※ Corresponding Author : Ihn-Han Bae, Address : (712-702) 330 Geumnak, Hayang, Gyeongsan, Gyeongbuk, Republic of Korea, TEL : +82-53-850-2742, FAX : +82-53-850-2740, E-mail : ihbae@cu.ac.kr
Receipt date : Mar. 5, 2007, Approval date : Jun. 27, 2007
† School of Computer and Information Eng., Catholic University of Daegu

tivities are effectively detected by using a rough membership function with the age of user profile. Our scheme uses the traveled cell ID, the duration of the service and the service class of a user as the feature value. When an intrusion occurs, the attacker masquerading the legitimate user trends to have a different used pattern. Therefore, we can detect anomaly by comparing the used patterns.

The rest of this paper is organized as follows. Section 2 gives a brief description of related works for intrusion detection techniques in mobile networks. Section 3 describes rough sets. Section 4 presents a rough set-based technique considering the age of user profiles for anomaly detection in mobile networks. Section 5 presents the simulation study of our proposed detection approach. In Section 6, we conclude this paper and point out future work.

2. RELATED WORKS

The number of intrusions into computer system is growing because new automated intrusion tools are appearing every day. Although there are many authentication protocols in cellular mobile networks, security is still a very challenging task due to the open radio transmission environment and the physical vulnerability of mobile devices.

Generally, there are two intrusion detection techniques, misuse based detection and anomaly-based detection. A misuse-based technique encodes the known attack signatures and system vulnerabilities. An anomaly-based detection technique creates normal profiles of system states and user behaviors and compares then against current activities.

Anomaly detection techniques assume that all intrusive activities are necessarily anomalous. This means that if we could establish a "normal activity profile" for a system, we could, in theory, flag all system states varying from the established profile by statistically significant amounts as intrusion attempts. However, if we consider that the set of intrusive activities only intersects the set of anomalous activities instead of being exactly the same, we find a couple of interesting possibilities: (1) Anomalous activities that are not intrusive are flagged as intrusive. (2) Intrusive activities that are not anomalous result in false negatives (events are not flagged intrusive, though they actually are). This is a dangerous problem, and is far more serious than the problem of false positives.

The main issues in anomaly detection systems thus become the selection of threshold levels so that neither of the above 2 problems is unreasonably magnified, and the selection of features to monitor. Anomaly detection systems are also computationally expensive because of the overhead of keeping track of, and possibly updating several system profile metrics. A block diagram of a typical anomaly detection system is shown in Fig. 1 [2,3].

The concept behind misuse detection schemes is that there are ways to represent attacks in the form of a pattern or a signature so that even variations of the same attack can be detected. This means that these systems are not unlike virus detection systems - they can detect many or all known attack patterns, but they are of little use for as yet unknown attack methods. An interesting point to note is that anomaly detection systems try to detect the complement of "bad" behavior. Misuse detection systems try to recognize known "bad" behavior. The main issues in misuse detection systems are how to write a signature that encompasses all possible variations of the pertinent attack, and how to write signatures that do not also

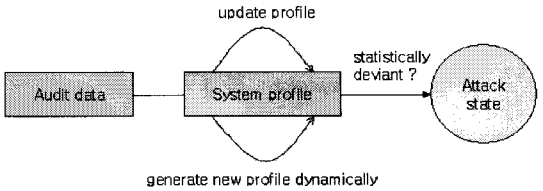


Fig. 1. A typical anomaly detection system.

match non-intrusive activity. A block diagram of a typical misuse detection system is shown in Fig. 2[2,3].

Y. Zhang proposed new model for Intrusion Detection System (IDS) and response in mobile, ad-hoc wireless networks. Each IDS agent runs independently and monitors local activities. It detects intrusion from local traces and initiates response. If anomaly is detected in the local data, or if the evidence is inconclusive and a broader search is warranted, neighboring IDS agents will cooperatively participate in global intrusion detection actions [1]. B. Sun proposed a mobility-based anomaly detection scheme in cellular mobile networks. The scheme uses cell IDs traversed by a user as the feature value [2]. O. Kachirski proposed multi-sensor intrusion detection system employing cooperative detection algorithm [4]. By efficiently merging audit data from packet level, user level and system level sensors, an entire ad hoc wireless network for intrusion is analyzed [5]. J. Gomez proposed a technique to generate fuzzy classifiers using genetic algorithms that can detect anomalies and some specific intrusions [6]. In [7], based on fuzzy view of rough sets instead of exact rules, a soft computing approach that uses fuzzy rules for anomaly detection is proposed. I. H. Bae proposed a rough set based anomaly detection scheme without considering the age of user profiles [8]. H. Deng proposed an agent-based cooperative anomaly detection methodology for wireless ad hoc networks [9]. The approach addresses the underlying distributed and cooperative nature of wireless ad hoc networks and adds one more dimension of cooperation to the intrusion detection process. That is, the anomaly detection is performed in a cooperative way involving the participation of multiple mobile nodes.

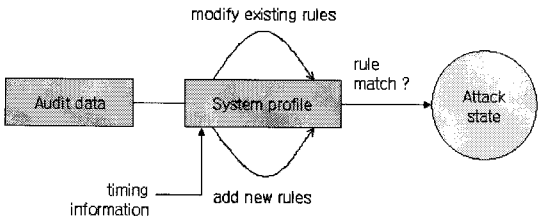


Fig. 2. A typical misuse detection system.

3. ROUGH SETS

The rough set is the approximation of a vague concept by a pair of precise concept, called lower and upper approximation, which are classification of domain of interest into disjoint categories. The rough set approach to processing of incomplete data is based on these approximations [10,11].

Let U be a finite set of objects called Universe, and $R \subseteq U \times U$ be an equivalence relation on U. The pair A=(U, R) is called approximation space, and equivalence classes of the relation R are called elementary sets in A.

For $x \in U$, let $[x]_R$ denote the equivalence class of R, containing x. For each $X \subseteq U$, X is characterized in A by a pair of sets - its lower and upper approximation in A, defined as:

$$\underline{A}X = \{x \in U \mid [x]_R \subseteq X\}$$
$$\overline{A}X = \{x \in U \mid [x]_R \cap X \neq \varnothing\}$$

The objects in $\underline{A}X$ can be with certainty classified as members of X on the basis of knowledge in R, while the objects in $\overline{A}X$ can be only classified as possible members of X on the basis of knowledge in A. The set $BN_A X = \overline{A}X - \underline{A}X$ is called the A-boundary region of X, and thus consists of those objects that we cannot decisively classify into X on the basis of knowledge in A.

Rough set can be also characterized numerically by the following coefficient called the accuracy of approximation, where *Card* denotes the cardinality.

$$\alpha_A(X) = \frac{Card\underline{A}X}{Card\overline{A}X}$$

Obviously $0 \leq \alpha_A(X) \leq 1$. If $\alpha_A(X)=1$, X is crisp with respect to A, and otherwise, if $\alpha_A(X) < 1$, X

is rough with respect to A.

Of course some other measures can also be defined in order to express the degree of exactness of the set X. It is possible to use a variety of $a_A(X)$ defined as

$$\rho_A(X)=1-\alpha_A(X)$$

and referred to as a A-roughness of X. Roughness as opposed to accuracy represents the degree of incompleteness to knowledge A about the set X [11].

4. ANOMALY DETECTION CONSIDERING THE AGE OF USER PROFILES

In this section, we introduce the construction of a rough set-based anomaly detection method that considers the age of user profiles. For each mobile user, the features of the user activities from the wireless application layer are captured. We use the traveled cell ID, the duration of the service and the service class of a user as the feature value. The feature values are stored in the user's feature profile database. In a cellular mobile network, this feature profile database is stored together with the mobile user's personal information, such as billing information, in the Home Location Register (HLR). We assume that HLR is secure and the feature information is accurate. Usually, because of its importance, HLR is protected with highly secure measure, and thus it is extremely hard to attack HLR.

The equivalence classes from the user's feature profile database are computed by using rough sets. For a mobile user, based on both the user activity information and the equivalence class information, a deviation number is computed by a roughness membership function with the age of the user profile, where the deviation number represents the degree that the user behavior is deviated from the normal behavior. When a user activity occurs, if the deviation number is greater than the deviation threshold that is a system parameter, an alert message occurs, otherwise the user activity identified as normal. The whole scheme is illustrated in Fig. 3.

A relation/view instance is a snap shot of a relational database, which represents user's instant perception of entities or objects represented in the database. An information system is such an instance. We should note that the information system is an extension of relational databases without the entity integrity constraint.

The user feature profile database used in our scheme is shown in Table 1, where REQ#, CELL, DUR, CLASS and AGE represent the service request number, the traveled cell ID, the requested service duration, the requested service class and

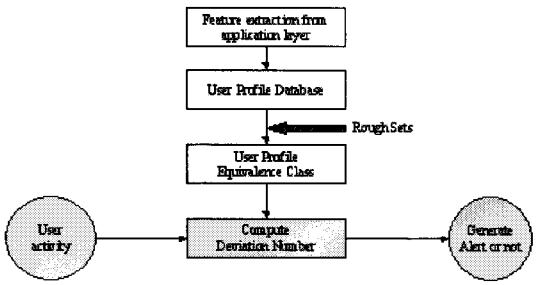


Fig. 3. The structure of rough set based anomaly detection scheme.

Table 1. User feature profile information system

REQ#	CELL	DUR	CLASS	AGE
1	a	1	α	3
2	a	1	α	3
3	a	2	α	3
4	b	2	β	3
5	b	2	β	3
6	b	2	β	2
7	b	2	β	2
8	b	3	γ	2
9	c	3	γ	2
10	c	3	γ	2
11	a	1	α	1
12	a	1	α	1
13	a	2	α	1
14	b	2	β	1
15	b	3	β	1

the age of the data, respectively. As the age of profile data is smaller, the profile data is newer. The profile data are continuously come and faded (aging) away.

In Table 1, let the attribute CLASS is decision attribute. Then we have the following three equivalence classes, called decision classes.

DE1={1, 2, 3, 11, 12, 13}={α}

DE2={4, 5, 6, 7, 14, 15}={β}

DE3={8, 9, 10}={γ}

For the conditional attributes (CELL, DUR), we have the following five equivalence classes, called condition classes.

CE1={1, 2, 11, 12}

CE2={3, 13}

CE3={4, 5, 6, 7, 14}

CE4={8, 15}

CE5={9, 10}

Comparing condition and decision classes, we get the inclusions.

CE1 ⊆ DE1

CE2 ⊆ DE1

CE3 ⊆ DE2

CE5 ⊆ DE3

The inclusion relation between CE4 and DE2 can be represented by a fuzzy inclusion. The fuzzy inclusions are represented by the inequalities of membership functions. Further, we will allow certain errors as long as they are within the radius of tolerance. The fuzzy inclusion is computed by Eq. (1), roughness membership function.

$$\mu(X,Y)=\frac{\sum_{age=1}^{g}\left[Card((\bar{R}X\cup\bar{R}Y)-(\bar{R}X\cap\bar{R}Y))\times(1-w_{age})\right]}{Card(\bar{R}X\cup\bar{R}Y)} \quad (1)$$

where g represents the number of age grades, w_{age} represents the weighted value of the age, and X and Y represent condition attribute and decision attribute, respectively.

For CE4 and DE2 those are not in inclusions, let X=CE4={b, 3}, Y=DE2={β}, w_1=0.6, w_2=0.3 and w_3=0.1. $\bar{R}X$={4,5,6,7,8,9,10,14,15}, and $\bar{R}X$={4,5,6,7,14,15}, The computed (X, Y)-roughness, $\mu(X,Y)=0.9/9\approx0.1$ by Eq. (1), and so that CE4 and DE2 is in 0.1-fuzzy inclusion.

$$CE4\subseteq_{(0.1)} DE2$$

We assume that the deviation threshold (ε) is 0.58. In case that a user activity (a, 3, α) is occurred, let X={a, 3} and Y={α}, $\bar{R}X$={1,2,3,7,8,9,10} and $\bar{R}Y$={1,2,3,11,12,13}, so that $\mu(X, Y)=4/10=0.4$. Accordingly, the user activity is evaluated as normal because that $\mu(X, Y)\leq\varepsilon$. But, in another case that a user activity (c, 2, α) is occurred, let X={c, 2} and Y={α}, $\bar{R}X$={3,4,5,6,7,8,10,14,15} and $\bar{R}Y$={1,2,3,11,12,13}, so that $\mu(X, Y)=8.4/14=0.6$. Accordingly, the user activity is identified as anomalous because that $\mu(X, Y)>\varepsilon$, and an alert message is generated.

5. PERFORMANCE EVALUATION

We use the following two metrics to evaluate the performance of our proposed anomaly detection scheme:

- Detection Rate: It is measured over abnormal itineraries. Suppose m abnormal itineraries are measured, and n of them are detected, detection rate in defined as n/m.
- False Alarm Rate: It is measured over normal itineraries. Suppose m normal itineraries are measured, and n of them are identified as abnormal, false alarm rate is defined as n/m.

We present and analyze the simulation results at different deviation threshold. In the simulation, we assume as follows. In cases of age 1 and age 2, if a user activity had two records those the two values of the user activity attribute values were matched with the feature values of the user profile data, the user activity is normal. Also, if a user activity had one record that two values of the user activity attribute values were matched with the

feature values of age 1 within the user profile data and one or more records those two values of the user activity attributes values were matched with the feature values of age 2 or age 3 within the user profile data, the user activity is normal. All user activities of other cases are abnormal. Table 2 shows the parameter values for the simulation.

The performance of the proposed scheme is compared with that of a rough set based anomaly detection scheme without aging [8]. Simulation results of the detection rate and the false alarm rate over deviation threshold are illustrated in Fig. 4. The detection rate of our scheme (*Detection (age)*) is better than that of the rough set based anomaly detection scheme without aging (*Detection (no age)*) regardless of deviation thresholds. While the false alarm rate of our scheme (*False Alarm (age)*) is nearly equal to that of the rough set based anomaly detection scheme without aging (*False Alarm (no age)*).

Table 2. Simulation parameters

Parameters	Values
The number of user activities	1,000
The traveled cell IDs	Random (1,4)
The type of service durations	Random (1,4)
The type of service classes	Random (1,3)
The weight of ages, (w_1, w_2, w_3)	(0.6, 0.3, 0.1)

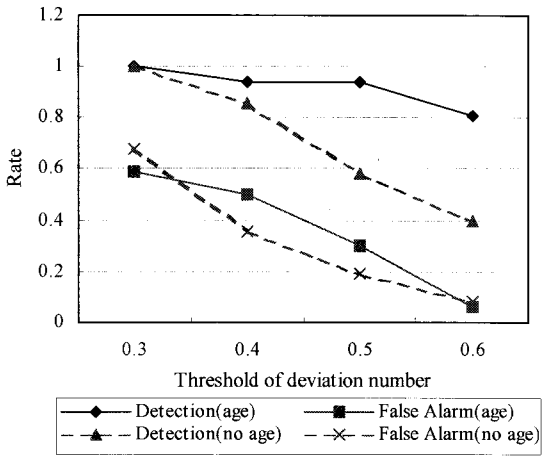


Fig. 4. Detection rate and false alarm rate at different deviation threshold.

Table 3 reports on numerical data for performance of anomaly detection schemes over deviation thresholds. From the results, when the deviation threshold is setting to 0.58~0.6, we can get the best performance of the proposed scheme that the detection rate is 0.878~0.813 and the false alarm rate is 0.146~0.069.

6. CONCLUSIONS

In this paper, we propose the intrusion detection technique that the normal profile of each mobile is constructed by using rough sets and anomaly activities are effectively detected by using a rough

Table 3. Numerical data for the performance of anomaly detection schemes over deviation thresholds

Deviation threshold	Aging		No Aging	
	Detection rate	False alarm rate	Detection rate	False alarm rate
0.3	1.0	0.596	1.0	0.687
0.35	1.0	0.566	1.0	0.565
0.4	0.943	0.506	0.866	0.363
0.45	0.943	0.417	0.866	0.363
0.5	0.943	0.306	0.585	0.196
0.55	0.878	0.205	0.585	0.196
0.58	0.878	0.146	0.415	0.083
0.6	0.813	0.069	0.415	0.083
0.65	0.46	0.04	0.324	0.063

membership function with the age of user profile. Our scheme uses the user activity information from the application layer as the feature value. Therefore, our scheme is used for all users because the feature values are selectively applicable to users. Simulation results demonstrate that the performance of our scheme depends on the deviation threshold. From the results, we know that the performance of our scheme is better than that of the rough set based anomaly detection scheme without aging regardless of deviation thresholds.

Future works include development of a new roughness membership function that reflects precise deviation number of user activities with ages and development of a rough set based anomaly detection scheme for mobile ad hoc networks.

REFERENCES

[1] Y. Zhang, W. Lee, and Y-A. Huang, "Intrusion Detection Techniques for Mobile Wireless Networks," *ACM/Kluwer Mobile Networks and Applications (MONET)*, Vol.9, No.3, pp. 545-556, 2002.

[2] B. Sun et al, "Mobility-Based Anomaly Detection in Cellular Mobile Networks," *WiSe'04*, pp. 61-69, 2004.

[3] A. Vattikonda et al, "Security in Mobile Computing Systems," *Department of Computer Science, The University of Kentucky, Term Paper*, 2003.

[4] Y. Zhang and W. Lee, "Intrusion Detection in Wireless Ad-Hoc Networks," *MobiCom'2000*, 275-283, 2000.

[5] O. Kachirski and R. Guha, "Effective Intrusion Detection Using Multiple Sensors In Wireless Ad Hoc Networks," *HICSS'03*, 2003.

[6] J. Gomez and D. Dasgupta, "Evolving Fuzzy Classfiers for Intrusion Detection," *Proceedings of the Workshop on Information Assurance United States Military Academy*, pp. 1150-1161, 2001.

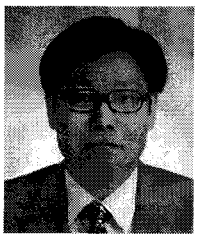
[7] T. Y. Lin, "Anomaly Detection - A Soft Computing Approach," *Proceedings of the 1994 Workshop on New security paradigms*, pp. 44-53, 1994.

[8] I. H. Bae, "Design and Evaluation of a Rough Set Based Anomaly Detection Scheme for Mobile Networks," *Advances in Natural Computation and Data Mining*, pp. 262-268, 2006.

[9] H. Deng, R. Xu, J. Li, F. Zhang, R. Levy, and W. Lee, "Agent-based Cooperative Anomaly Detection for Wireless Ad Hoc Networks," *Proceeding of the 12th International Conference on Parallel and Distributed Systems*, pp. 613-620, 2006.

[10] R. Jensen and Q. Shen, "Fuzzy-Rough Sets for Descriptive Dimensionality Reduction," *Proceedings of the 11th International Conference on Fuzzy Systems*, pp. 29-34, 2002.

[11] Z. Pawlak, *Rough Sets Theoretical Aspects of Reasoning about Data*, Kluwer Academic Pub., pp. 252, 1991.



Ihn-Han Bae

Received the BS degree in Computer Science from Kyungnam University, Korea, in 1984, and the MS and the Ph.D. degrees in Computer Engineering from Chungang University, Korea in 1986 and 1990, respectively. From 1996-1997, he was a post-doctoral at the Department of Computer and Information Science in the Ohio State University, USA. From 2002-2003, he was a visiting professor at the Department of Computer Science of Old Dominion University, USA. Currently, he is a professor at the School of Computer and Information Communication Engineering in Catholic University of Daegu, Korea. His research interests include multimedia systems, wireless networks, wireless Internet and P2P systems.