IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661, ISBN: 2278-8727Volume 6, Issue 5 (Nov. - Dec. 2012), PP 42-47
www.iosrjournals.org

Data Mining For Intrusion Detection in Mobile Systems

[1,] Seyed Hasan Mortazavi Zarch, [2,] Farhad Jalilzadeh, [3,] Madihesadat Yazdanivaghef
[1,2,3] *Department Of Computer Science and System Engineering, Andhra University, India*

Abstract: *New security threats emerge against mobile devices as the devices' computing power and storage capabilities evolve. Preventive mechanisms like authentication, encryption alone are not sufficient to provide adequate security for a system. There is a definite need for Intrusion detection systems that will improve security and use fewer resources on the mobile phone. In this work we proposed an intrusion detection method that efficiently detects intrusions in mobile phones using Data Mining techniques. We used network based approach that will remove the overhead processing from the mobile phones. A neural network classifier will be built and trained for each user based on his call logs .An application that runs on smart phone of the user collects certain information of the user and sends them over to the remote server. These logs then fed to the already trained classifier which analyzes the logs and sends back the feedback to the smart phones whenever abnormalities are found. Also we compared different neural classifiers to identify the classifier with better performance. Our results showed clearly the effectiveness of our method to detect intrusions and outperformed existing Intrusion detection methods with 95% detection rate.*

I. Introduction

The advanced mobile communication devices such as smart phones are changing the way in which we communicate process and store data from any place. They evolved from simple mobile phones into sophisticated and yet compact mini computers [2]. They are not just voice communication devices. Apart from browsing internet these devices can receive e-mail send MMS messages, exchange information by connecting to other devices. They are also equipped with operating system, text editors, spreadsheet editors and database processors. As the capabilities of mobile devices evolve, their usage for processing and storing private and business data is likely to increase. Currently 200 million users are using smart phones worldwide and the number of people using smart phones will likely to increase to 1 billion in the next 2 years. That is approximately one sixth of the world population and equivalent to population of India.

As these devices can allow third party software's to run on them they are vulnerable to various threats like viruses, malware, worms and Trojan horses. Also a mobile device can initiate communication on anyone of its communication interfaces and also can connect to wide variety of wireless networks. Intrusion prevention mechanisms such as encryption, authentication alone cannot improve the security of the system. Already existing desktop based Intrusion Detection software's may not be good for mobile systems because of the memory consumption rate and power consumption. We need to come up with Intrusion detection systems that will not only improve the security of these systems but also reduce the processing overhead from the system.

1.2 Contribution of this work

In this work we focus on designing a good intrusion detection mechanism for Mobile phone systems using Data Mining techniques. Our method uses network based approach which significantly removes the overhead of processing from the mobile phone system and efficiently detects intrusions. We used Data Mining techniques in order to detect intrusions.

Our method has the following advantages

1. Reduces the processing overload from the mobile phone
2. Detects intrusions at high accuracy rate

1.3 Outline

The thesis is organized as follows: In section2 we will discuss about the different types of attacks on mobile phones, Intrusion Detection methods and techniques and related work on Intrusion detection in mobile phones. In section3 we will discuss our new proposed method. In section4 we discuss about the experiments and results. Last section concludes and explains the future work.

II. Background and Related Work

2.1 Types of attacks on Mobile phones

2.1.1 Malware:

Malware short form for malicious software is software designed to secretly access a system without the owner's informed consent. On a mobile device, a piece of malware can hinder or prevent the normal usage of the device by consuming computing, memory and I/O resources, or just make the device stop responding to user's actions ,reveal private information to unauthorized parties without the user even knowing that. It can destroy or modify data on the device, initiate unwanted communication: the owner of the device often has to pay for the data transmission. A piece of malware may, for example, make the device to place a call to some expensive service

2.1.2 Trojan:

The primary concern of malicious attack is from Trojan applications. Nowadays, it is very common to find a computer Trojan that transmits spam emails to Internet user. This will interrupt network performance and create lots of inconvenient issues to user, but generally involves no direct cost to the user. However, a similar Trojan on a phone could impose a heavy financial penalty on the consumer. For example [5], an application that sends messages at a rate of $1/sms, if the infected application runs for an average of 100 SMS/day, it would cost user $100/day. Consumers will only receive mobile billing after 30 days before the mobile owner realizes that his phone has been infected by malicious program.

2.1.3 Worm

The second area of attack is to develop a self-replicating mobile application. This type of malicious program can be developed from a Trojan by attaching a copy of itself to the MMS messages. For such viruses to work, interaction with the message recipient is required. But one thing for sure is that MMS message service does not allow any application file to attach with it for this moment

2.1.4 Virus

Virus is the most destructive program designed to damage files or otherwise interfere with the mobile phone's operation. Similar attacks can be developed for other nefarious reasons, such as copying the contents of the phone's address book and sending them elsewhere, corrupting or deleting the numbers in the address book, blocking incoming call, and a whole host of other denial of service attacks.

2.2 Intrusion Detection Methods

Any intrusion detection method is dependent on the type of monitoring data that is given as input to the detection algorithm, regardless of the type of the actual detection method used. Based on the type of monitoring data used intrusion detection methods can be classified as either host based or network based. Network based intrusion detection often performed on network data like network traffic, data packets, etc. Where as host based intrusion detection is performed on data available on the mobile phone like System level events, CPU activity, memory consumption, file I/O activity, and, network I/O activity, operating system level events, Application level, Measurements. Both of them have their advantages and disadvantages.

2.2.1 Network Based Monitoring

Advantages:

- Reduces the processing overload from the mobile phone
- Detects external intrusions

Disadvantages:

- No access to monitoring data on the mobile phone that is useful for detection
- Communication environment is very fragmented as a mobile device can be connected to multiple sources on multiple interfaces at the same time
- Cannot detect intrusions on the device itself like malware etc.
- Collecting all relevant network based monitoring data for all networks and communication interfaces that the mobile device interacts with may be very difficult

2.2.3 Host Based

Advantages:

 Have access to private information on the mobile device that is useful to detect intrusions as the information collected from the mobile device will reflect the device behavior accurately

- intrusion detection models with host-based data collection provide more accurate and reliable results than other approaches[4]

Disadvantages:

- Difficult to implement because of the processing limitations of mobile phones
- Difficult to provide security for the data that is directly collected from the mobile device

2.4 Intrusion Detection Techniques

2.3.1 Anomaly Detection:

The goal of anomaly detection is to identify cases that are unusual. Anomaly detection is an important tool for detecting fraud, network intrusion, and other rare events that are significant but hard to find.

The main advantage of anomaly detection is that it does not require prior knowledge of intrusion and can thus detect new intrusions. The main disadvantage is that it may not be able to describe what the attack is and may have high false positive rate.

I. Misuse Detection

Misuse detection systems use patterns of well-known attacks or weak spots of the system to match and identify known intrusions. In misuse detection, the IDS analyzes the information it gathers and compares it to large databases of attack signatures. Essentially, the IDS looks for a specific attack that has already been documented.. For example, signatures rule for the "guessing password attack" can be "there are more than 4 failed login attempts within 2 minutes". The main advantage of misuse detection is that it can accurately and efficiently detect instances of known attacks. The main disadvantage is that it lacks the ability to detect the truly innovative (i.e., newly invented) attacks.

2.4 Related Work:

IDAMN: Samfat et.al [3] 1997 proposed Intrusion detection architecture for mobile networks. Their systems main functionality is its ability to perform intrusion detection in the visited location and within the duration of a typical call. They used both rule based and anomaly detection methods. It is effective in detecting intrusions like an user active in 2 different locations at the same time, An area having low network activity experiencing high network activity and also effective in detecting anomalous behavior. It has more than 70% intrusion detection rate.

Mirela et.al[1] proposed an neural fraud system to identify frauds and impostors and improper use of mobile phones using neural network classifier. Their method takes the current and the past activities of the user and classifies the users into groups. And as soon as an intrusion or fraud is detected the user will be notified about the same.

Jerry et.al proposed SmartSiren [9], a Virus detection and alert system for Smartphone's. Their main approach was to detect the viruses that are transmitting through communicating interfaces like SMS, Bluetooth, and Infrared by monitoring the communication activity of the mobile phone. The SmartSiren architecture consists of a back end server that communicates with light weight agents on the protected devices. The agent collects the information and sends them to the backend proxy server to perform analysis and send out the alerts.

III. Proposed Intrusion Detection Method

3.1 Drawbacks of the Existing Methods

Although significant amount of research has been done and many methods proposed on detecting intrusions in mobile phones there are still some drawbacks in the existing methods.

The Neural network approach proposed by Mirela et.al is effective in detecting fraud calls and imposters. The disadvantage of this method is that the process is relatively slow and this method concentrated classifies users into groups having same behavior and hence there will be lot of false positives.

Intrusion Detection Architecture for Mobile Networks (IDAMN) proposed in 1997 is also effective in detecting intrusions but the disadvantage of this method is that it can only be applied to specific network architectures using specific hardware.

Smart Siren developed by jerry et.al focused on viruses that are transmitted through SMS messages and other communication interfaces like Bluetooth and Infrared. But they did not concentrate on worms that will automatically make high rate calls from the mobile which will incur loss to the user as they are only collecting and monitoring SMS traces.

Thus the drawbacks of the existing intrusion detection methods are firstly, they are host based meaning that they will run on the mobile phones draining the power resources and also consuming the CPU power. Secondly, some of the existing methods are rule based that means they will detect only the known intrusion and will fail to detect some innovative attacks which do not have an existing signature rule.

Thirdly the methods will not be valid for rapidly changing mobile technologies. A technique that was developed by taking the network architecture and mobile phone specifications into consideration may not be efficient because each mobile may use its own operating system and each mobile carrier network may use its

own network architecture.

By considering all these drawbacks we came up with a network based Intrusion detection method that reduces processing overhead from the mobile phone and works irrespective of the type of operating system on the mobile phone and the network architecture.

3.2 Design and Detection Process

We have developed a new Data Mining approach for detecting intrusion in mobile phone systems. As we discussed in our previous sections there are certain type of worms that will make calls from the mobile phone without the user knowledge. Those types of worms may also send text messages/make calls involving money transfer requests without the user knowledge. It is very difficult to detect those attacks and the user may know about this attack only when he/she sees her bill at the end of the month. This will incur a loss to the customer as well as the telecom provider. Our approach will be useful in detecting those types of attacks and will send an alert immediately to the user instead of waiting until the last day of the month.

We basically collect the call logs of the user who wants to use this service and we will build a classifier for that particular user. We will train the classifier using the call logs of the user. An application will be installed on the user system. Every day at a specified time this app on his smart phone will upload the call logs to the remote server. The uploaded call logs are then processed and will be fed to the already trained classifier. The already trained classifier will have the prior knowledge of the user normal behavior. The classifier analyzes these call logs and if the classifier identifies any abnormalities then it will send an e-mail/ Text message to the user based upon his choice. To reduce the false alarms we will send an alarm to the user only if the observed abnormalities are more than a specific threshold value.

We used Network based approach. As the application which collects phone logs will reside on the mobile and sends the logs to remote server (Network Based). This will remove the overhead of processing from the mobile phone which saves the limited battery power and the computing power for other purposes.

We compared Naïve-Bayes neural network classifier and SVM [7] Classifier to classify the call logs to see which classifier gives optimal performance. Weka, a Data Mining tool is used for the purpose of simulations. We collected call logs of 4 users of different sizes to evaluate the size of the dataset on the performance. We used Ten-Fold cross validation to measure the performance of the classifier.

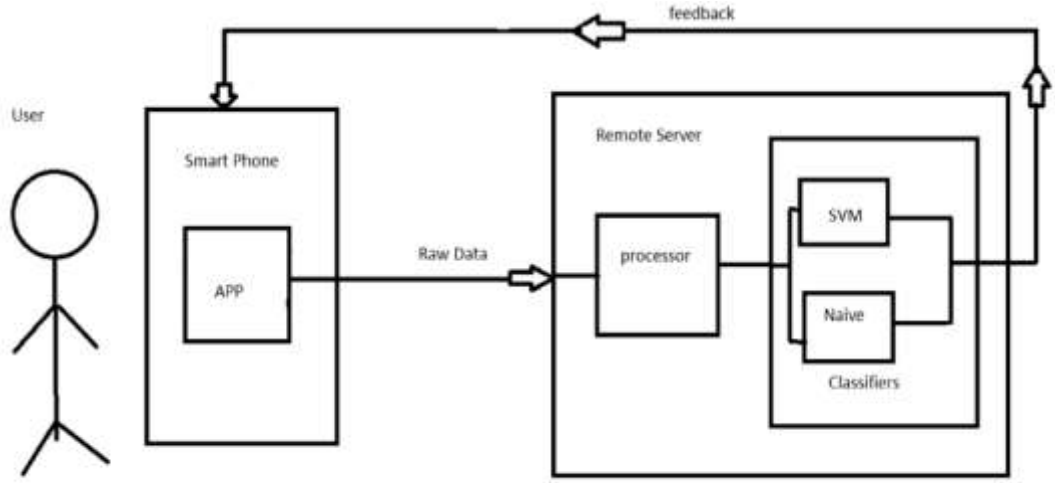

Figure 1: Architecture of the proposed system

IV. Simulation Results

4.1 Experiment Setup

The goal of our simulations is to show how neural network classifiers can efficiently and effectively able to detect intrusions. We choose two neural network classifiers Naïve-Bayes and Support Vector Machine (SVM). We carried out the experiments using Weka, a Data Mining tool. The data files (call logs) are collected from friends and relatives. Next we will discuss about the data used to train and test the classifiers.

Call Logs:

There will be many attributes when considering phone logs but we considered the following 4 attributes that we think will be sufficient and useful for the classification

1. Day of the week (Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday)
2. Duration of the call (minutes)
3. Type of the call (Local, Long Distance, International)
4. Time of the call (Morning, Afternoon, Evening, Night, Mid-Night)

The attribute "Day of the week" indicates the particular day of the week when the call was made, "Duration of the call" gives you the length of the call in minutes, "Type of call2 " tells you whether a call is local/Long Distance/International, "Time of the call" is used to find at what time of the day a call was made. The raw data that is collected from a mobile phone will be in the following format.

Example log files (Before preprocessing)

Date	*Destination*	*Time*	*Number*	*Call*	*Minutes*
8/27/09	*Columbia, MO*	*11:11 AM*	*573-529-9284*		*1*
8/27/09	*Ames, IA*	*6:33 PM*	*515-598-5499*		*2*
8/28/09	*Incoming*	*9:23 AM*	*573-529-9284*		*2*
8/28/09	*Columbia, MO 2:53 PM*		*573-529-9284*		*1*
8/28/09	*Incoming*	*3:27 PM*	*573-529-9284*		*2*
8/28/09	*Columbia, MO 6:26 PM*		*573-529-9284*		*1*
8/28/09	*Columbia, MO 6:46 PM*		*573-529-9284*		*1*
8/28/09	*Columbia, MO 6:49 PM*		*573-529-9284*		*1*
8/28/09	*Columbia, MO 7:35 PM*		*573-529-9284*		*1*
8/29/09	*Incoming*	*1:01 PM*	*573-529-9284*		*2*
8/29/09	*Incoming*	*1:07 PM*	*573-529-9284*		*1*
8/29/09	*Ames, IA*	*8:36 PM*	*515-292-2077*		*2*

These data is classified in such a way that the date is transformed into Day, destination into "Type of call1", Time into "Time of call 2", Minutes into "duration of call". Following is an example of such formatted data.

Processed Data Files (.arff files):

Weka recognizes the data files which are in .ARFF format. Below is the sample .ARFF data file

@relation logs-simplified
@attribute day {Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday} @attribute minutes numeric
@attribute typeofcall2 {LO, LD, IL} @attribute time {AM, PM, EV, NI, MN} @attribute class {no, yes}
@data
Thursday, 1, AM, LD, NO
Thursday, 16, EV, LO, NO
Thursday 4, AM, LO, NO
Thursday, 2, PM, LD, NO
Friday, 2, PM, LO, NO
Friday, 1, EV, LD, NO
Friday, 2, EV, LD, NO
Friday, 2, EV, LD, NO
Friday, 1, EV, LD, NO
Friday, 2, PM, LO, NO
Friday, 1, PM, LO, NO
Friday, 1, PM, LD, NO

We then add some known fraud data into these call logs build models for each user using Naïve-Bayes and SVM. Both Naïve Bayes classifier and SVM are used for classification to compare which classifier yields better performance. We also added some known fraud data into the call logs when training the classifier. After doing several experiments on both SVM and Naïve-Bayes we got the following results.

10- Fold cross validation

- 10-fold cross validation is used in the field of machine learning to determine how accurately a learning algorithm will be able to predict data that it was not trained on

- The training dataset is randomly partitioned into *10* groups, first 9 groups are used for training the classifier and the other group is used as testing dataset.
- process is repeated until all groups are covered and the performance is measured as the aggregate of all the 10 folds

4.2 Results

In our experiments the outcomes are labeled either as positive (YES) or negative (NO) class. There are four possible outcomes from the classifier. If the outcome from a prediction is YES and the actual value is also YES, then it is called a true positive (TP); however if the actual value is NO then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are NO, and false negative (FN) is when the prediction outcome is NO while the actual value is YES.

To draw an ROC curve [11], only the true positive rate (TPR) and false positive rate (FPR) are needed. TPR determines a classifier or a diagnostic test performance on classifying positive instances correctly among all positive samples available during the test. FPR defines how many incorrect positive results occur among all negative samples available during the test. FPR and TPR are calculated by using these formulae.

$$TPR = TP / (TP + FN)$$
$$FPR = FP / (FP + TN)$$

The best possible prediction method would yield a point in the upper left corner or coordinate (0, 1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (false positives). ROC curve is plotted for each user to analyze the performance of the classifier.

V. Conclusion And Future Work

In this work we presented an intrusion detection method for the smart phones using Data mining techniques followed by its performance using real time user data. We analyzed the performance of two classifiers Naïve-Bayes and SVM. Both of the classifiers are very effective in detecting intrusions. Both of them have shown a performance rate higher than 90%. However SVM 2 class classifier performance is slightly better than that of Naïve-Bayes. We also analyzed the effect of size of the dataset on the performance and our method performed well irrespective of the size of the data set used. As we used network based approach in our method, it significantly reduced the processing overhead from the mobile phones.

Our results showed clearly the effectiveness of our method to detect intrusions. The Detection rate of our method outperformed previously proposed intrusion detection methods. We compared our method with the existing Intrusion detection methods like Intrusion Detection Architecture for Mobile Networks (IDAMN). Our method clearly outperformed the IDAMN with a detection rate of more than 95% and false alarm lower than 4 % compared to IDAMN which has 70% detection rate and 5% false alarm rate. Thus our method proved to be efficient in detecting intrusions.

As a future work we plan to extend this work for features like location of the user,
Email logs, text logs, operating system level events etc. for monitoring purposes. Also we would like to develop a User Interface where the user can easily access his past records, Intrusion alerts he received and view his records anytime.

Bibliography

[1] Azzedine Boukerche, Mirela Sechi M, Annoni Notare "Behavior-Based Intrusion Detection in Mobile Phone Systems", Journal of Parallel and Distributed Computing 62, pp.1476–1490, 2002.
[2] Shabtai, A., et al. Intrusion detection for mobile devices using the knowledge-based, temporal abstraction method. J. Syst.Software (2010), doi:10.1016/j.jss.2010.03.046
[3] Didier Samfat, Refik Molva, "IDAMN: an Intrusion Detection Architecture for Mobile Networks", INSTITUT EURÉCOM, France
[4] Markus Miettinen, Perttu Halonen, Kimmo Hatonen, "Host-Based Intrusion Detection for Advanced Mobile Devices", Nokia Research Center, IEEE, 2006.
[5] S. Shimojo et al. (Eds.): HSI 2005, LNCS 3597, pp. 57 – 65, 2005. Springer-Verlag Berlin Heidelberg 2005
[6] Wenkee Lee, "Intrusion Detection Techniques for Mobile Wireless Networks", Mobile Networks and Applications, 2003.
[7] Eugene spafford, Diego Zamboni, "Data collection mechanisms for intrusion detection systems", CERIAS Technical Report, West Lafayette, IN, 2000.
[8] Jerry Cheng, Starsky H.Y.Wong, Hao Yang and Songwu Lu, "SmartSiren: Virus Detection and Alert for Smartphones", MobiSys'07, San Juan, Puerto Rico, 2007.
[9] Barak Pearlmutter, Christina Warrender, Stephanie Forrest, "Detecting Intrusions Using System Calls:Alternative Data Models", New Mexico., 2010.