TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 16, No. 2, November 2015, pp. 350 ~ 354
DOI: 10.11591/telkomnika.v16i2.8441

Machine Learning Based Automotive Forensic Analysis for Mobile Applications Using Data Mining

MD. Hussain Khan*, G. Pradeepini
Dept of CSE, K L University, Vaddeswaram, Guntur, A.P, India
*Corresponding author, e-mail: khans.90786@gmail.com

Abstract

Phone is a device which provides communication between the people through voice, text, video etc. Now a day's people may leave without food but not without using phones. No of operating systems are working with various versions and various security issues are working. Security is very important task in Mobiles and mobile apps. To improve the security status of mobiles, existing methodology is using cloud computing and data mining. Out traditional method is named as MobSafe to identify the mobile apps antagonism or graciousness. In the proposed system, we adopt Android Security Evaluation Framework (ASEF) and Static Android Analysis Framework (SAAF).In this paper, our proposed system works on machine learning to conduct automotive forensic analysis of mobile apps based on the generated multifaceted data in this stage.

Keywords: *ASEF, SAAF, OS, mobiles*

1. Introduction

Android could be mobile software (OS) taking into consideration the Linux half and at the moment grew by Google. With a shopper interface taking into consideration direct management, robot is made public essentially for bit screen cell phones, for instance, PDAs and pill PCs, with specific shopper interfaces for TVs (Android TV), autos (Android Auto), and wrist watches (Android Wear). The OS uses bit inputs that freely compare to real activities, like swiping, tapping, squeezing, and converse compressing to manage on-screen articles, and a virtual console.

Other than empowering correspondence with the cloud, Android's remote APIs in addition empower correspondence with totally different gadgets on constant neighborhood system, and even gadgets that aren't on a system, however rather square measures physically close-by. The enlargement of Network Service Discovery (NSD) takes this any by allowing associate degree application to go looking out associate degree adjacent contrivance running administrations with that it will impart. Coordinating this utility into your application helps you provide an in depth type of highlights, for instance, taking part in amusements with purchasers within the same area, propulsion footage from associate degree organized NSD-empowered digital camera, or remotely work into totally different machines on constant system.

1. Similar to past studies, we found wide misuse of privacy sensitive information—particularly phone identifiers and geographic location. Phone identifiers, e.g., IMEI, IMSI, and ICC-ID, were used for everything from "cookie-esque" tracking to accounts numbers.

2. We found no evidence of telephony misuse, background recording of audio or video, abusive connections, or harvesting lists of installed applications.

3. Ad and analytic network libraries are integrated with 51% of the applications studied, with Ad Mob (appearing in 29.09% of apps) and Google Ads (appearing in 18.72% of apps) dominating. Many applications include more than one ad library.

4. Many developers fail to securely use Android APIs. These failures generally fall into the classification of insufficient protection of privacy sensitive information. However, we found no exploitable vulnerabilities that can lead malicious control of the phone.

Received July 2, 2015; Revised October 4, 2015; Accepted October 19, 2015

2. Related Work

Mobile phone security becomes very important in real world. Some of the security issues in android applications are:

Enck et al. [13] describe the design and usage of a structure to tell apart conceivably vindictive applications in light-weight of authorizations asked for by Android applications. The structure peruses the pronounced consents of Associate in Nursing application at introduce time and analyzes it against an appointment of principles honored to talk to risky conduct. for example, Associate in Nursing application that demands access to poring over phone state, record sound from the electronic equipment, and access to the net may send recorded phone discussions to a foreign space. The system empowers applications that do not announce (known) unsafe authorization mixes to be introduced consequently, and concedes the approval to introduce applications that do to the shopper.

Ontang et al. [18] gift a fine-grained access management strategy foundation for guaranteeing applications. Their proposition expands the current golem consent show by allowing authorization articulations to precise additional detail. for example, rather than simply allowing Associate in Nursing application to send IPC messages to a different taking into consideration consent names, association is more to point wants for setups or programming variations. The creators highlight that there are true utilize cases for a additional complex strategy accent, particularly in light-weight of the actual fact that untrusted outsider applications typically collaborate on golem.

On the theme of examination of consent based mostly architectures, Barth et al. [10] investigated twenty five program expansions for Firefox and recognized that seventy eight are given a bigger variety of advantages than ought to be expected, increasing the assault surface on these highlight upgrading extra things. The investigation lead the creators to the configuration of Associate in Nursing authorization based mostly framework for program expansions in Google Chrome. The framework controls access to bookmarks, tabs, and areas accessible to a selected growth. Researching the convenience of use of authorization based mostly architectures, Reeder et al. [19] engineered up a structure for showing and sterilization document consents on a Winows operating framework. They used a grid based mostly perception referred to as expandable lattice, which provides a theoretical illustration of document consents in a very graphical arrangement. Their shopper studies incontestable this matrix perception permits purchasers to complete errands speedily and every one the additional exactly.

Amandroid may be a static examination system for Android applications. The Android stage is massively current. Be that because it could, pernicious or defenseless applications are accounted for to cause a couple of security problems. As of currently there's no powerful technique that a business administrator will use to vet applications getting into a business (e.g., Google Play).

Former works utilizing static investigation to deal with Android application security problems additional think about specific problems and made specific instruments for them. we tend to watch that a large a part of those security problems is determined by tending to at least one hidden center issue – catching linguistics practices of the applying, for instance, item focuses to and control-/information stream knowledge. after, we tend to made public otherwise to subsume leading static investigation for validatory Android applications, and made a bland structure, referred to as Amandroid, that will stream and association delicate data stream examination in a very between phase manner.

Our methodology demonstrates that a whole static investigation technique on Android applications is totally sensible relating to reckoning assets, and therefore the Amandroid structure is all-mains and straightforward to be extended for a few forms of specific security examinations.

Since Amandroid squarely handles Inter-segment management and data streams, it is used to deal with security problems that outcome from communications among various elements from either an equivalent or numerous applications. Amandroid investigation is sound therein it will provide certification of the unfortunate deficiency of the predefined security problems in Associate in Nursing application with all around indicated and wise suspicions on the Android runtime and its library.

On prime of Amandroid we tend to performed sure specific security examinations, for instance, a) shopper secret word stream following b) set up infusion identification, and c) crypto

API abuse checking. we tend to apply those examinations on many applications gathered from Google Play's illustrious applications Associate in Nursingd an outsider security organization, and therefore the outcomes demonstrate that it's appropriate discovering real security problems and sufficiently effective as so much as examination time.

3. Security in Android Apps

Android has security highlights incorporated with the package that altogether decrease the return and impact of utilization security problems. The system is planned therefore you'll unremarkably manufacture your applications with default framework and document consents and keep from difficult decisions regarding security.

A share of the middle security includes that assist you fabricate secure applications include:

1) The automaton Application Sandbox,that detaches your application data and code execution from completely different applications.
2) An application structure with powerful usage of normal security utility, for instance, cryptography, consents, and secure IPC.
3) Technologies like ASLR, NX, ProPolice, safe_iop, OpenBSD dlmalloc, OpenBSD calloc, and Linux mmap_min_addr to moderate dangers connected with basic memory administration mistakes.
4) An disorganized record framework that may be authorized to confirm data on lost or purloined gadgets.
5) User-allowed authorizations to confine access to framework highlights and consumer data.
6) Application-characterized consents to manage application data on AN each application.

4. Existing MOBSAFE

Saturn-cloud a home-brewed cloud computing platform is used to conduct security analysis task. Saturn-storage, NFS storage with ZFS file system (open indiana+napp-it), is used to accommodate the virtual machines. It can scale to 16 hard disks, each with 2 TB SATA storage, totally achieve 32 TB store volume. Cloudstack is used to manage a VMware vSphere based computing servers.

4.1. Hadoop Storage for Mobile Apps

There are about 40 servers and 40 TB storage in our experimental research platform based on HDFS.

4.2. ASEF

ASEF is Associate in automatize tool which may be utilized to dissect automaton application. after you gift Associate in Nursing obscure apk file to ASEF for investigation, firstly it'll begin the ADB work and traffic, sniffing utilizing TCPDUMP, then dispatch Associate in Nursing automaton Virtual Machine (AVD) and introduce the applying thereon. at that time ASEF starts to dispatch the applying to be cleft and send varied irregular motions to reenact human incorporation on the applying. Within the mean solar time, ASEF likewise cares the log of automaton virtual machine with a CVE library, and its net action with Google Safe program API. Once a selected range of signals area unit sent to virtual machine, the check circle is finished and also the application are uninstalled. At that time ASEF can begin to look at the log file and also the net traffic that the applying created. ASEF uses Google Safe Browsing API to find out whether or not the URLs the applying conceives to reach area unit harmful or not. ASEF in addition checks the existed unprotectedness with a noted quality summation to find out whether or not the applying encompasses a few real weaknesses.

4.3. SAAF

SAAF could be a static instrument for automaton apk files. It will disentangle the substance of apk files, and decipher the substance to smali code, then it'll apply program cutting on the smali code, to interrupt down the consents of applications, match heuristic examples, and perform system cutting field officer

5. Machine Learning

Machine Learning is a part of computer science. It is the study of pattern recognition and other learning algorithms. Machine learning targets on prediction, based on the familiar properties learned from the trained datasets. In this paper, our proposed system works on k-means method and it will classify the mobile apps from various sources and to know the apps are antagonism and graciousness.

5.1. K-Means

In this paper, our approach is K means clustering algorithm classifies the mobile apps antagonism or graciousness. K-means algorithm is a simple and most popular algorithm for clustering analysis. It is based on classification of the objects based on attributes or features into “K” number of group. ”K” is an Integer value.

5.2. Working of K-Means

It is very simple learning algorithm for any of clustering analysis. The aim of K-means algorithm is to find out the best classification of n entities in K groups. Actually the aim is to set the classification of n entities into k sets Si = 1,2,3…….K in order to decrease the distance between clusters.

$$\mathbf{J=}\ \sum_{j=1}^{k}\sum_{i=1}^{n} \| x_i^j - c_j \|^2$$

J= Objective Function.
K=No of Clusters.
n=no of cases.
x=case i
c=centroids for cluster j
and distance function.

5.3. Algorithm

1. Clusters the data into k groups where k is predefined.
2. Select k points at random as cluster centers.
3. Assign objects to their closest cluster center according to the Euclidean distance function.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

6. Experimental Results

It is the process of identifying set of similar objects called clusters. There may be number of clusters if the object is placed in one cluster it is similar in some sense.

“K” in K-means will give the number that will going to have to feed to our algorithm. “K” is defined as no of clusters to find out the data. Each Clusters will form the similar objects as one cluster.

“K” means divide the data into “K” distinct clusters. If we give the number of K=5 it will from 5 clusters and if the no of k=3, it form 3 clusters.

The grouping of objects done by the data collected from the various sources. The classification done based on the logging and network behavior data. According to the data from the dataset our proposed system works on trained dataset. To know the accuracy metrics includes precision and recall can be measured to evaluate the classifier algorithm.

From the confusion matrix:

	C1	C2
C1	True Positive (TP)	False Negative(FN)
C2	False Positive (FP)	True Negative(TN)

Recall=TP/ (TP+FN), also called positive predictive value. Is the fraction of retrieved instances that are relavant?
Precision= TP/ (TP+FP), is the fraction of relevant instances that are retrieved.

The K-means algorithm is become very fast and powerful which will use it on the dataset with no of dimensions. In this paper, K-means clustering will classify the mobile apps according to the antagonism and graciousness.

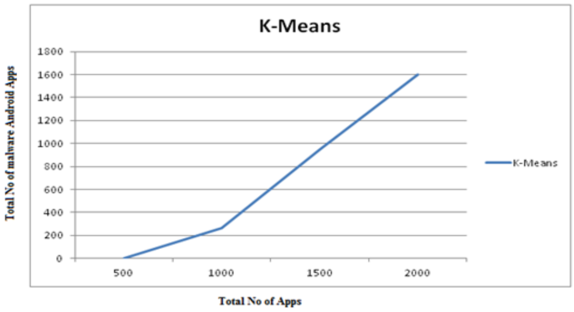

Figure 1. Finding the malware android apps by K-Means Algorithm

7. Conclusion

Machine Learning (ML) is a promising technology to identify mobile app's virulence or benignancy based on data mining. As we collect more and more app's logging and network behaviour data, we can further use K-means method to classify apps after training a classifier. In this case, the well-known accuracy metrics includes precision and recall can be measured to evaluate the calssifier algorithm.

References

[1] Jianlin Xu, Yifan Yu, Zhen Chen, Bin Cao, Wenyu Dong, Yu Guo, Junwei Cao. MobSafe: Cloud Computing Based Forensic Analysis for Massive Mobile Applications Using Data Mining.
[2] Greg Hamerly, Charles Elkan. Learning the k in k-means. Department of Computer Science and Engineering University of California, San Diego La Jolla, California.
[3] Ibrar Hussain. Clustering in Machine Learning.
[4] R Lawler, Mary Meeker's. Internet Trends report. 2013.
[5] J Wu. On Top of Tides, Beijing: China Publishing House of Electronics Industry. 2011.
[6] SQ Feng. Android software security and reversing engineering analysis. Beijing: Posts and Telecom Press. 2013.
[7] Gartner. 2012. http://www.gartner.com/it/page.jsp?id=2153215.
[8] List of mobile software distribution platforms. 2013. http://en.wikipedia.org/wiki/List of digital distribution platforms for mobile devices.
[9] D Barrera, HG Kayacik, PC van Oorschot, A Somayaji. *A methodology for empirical analysis of permission-based security models and its application to Android.* In Proc. 17th ACM Conference on Computer and Communications Security. Chicago, USA. 2010: 73- 84.
[10] W Enck, D Octeau, P McDaniel, S Chaudhuri. *A study of android application security*. In USENIX Security Symposium. San Francisco, USA. 2011.
[11] AP Felt, E Chin, S Hanna, D Song, D Wagner. *Android permissions demystified.* In Proc. 18th ACM Conference on Computer and Communications Security. Chicago, USA. 2011: 627-638.
[12] KO Elish, D Yao, BG Ryder. *User-centric dependence analysis for identifying malicious mobile apps*. In Workshop on Mobile Security Technologies (MoST). San Francisco, USA. 2012.
[13] I Burguera, U Zurutuza, S Nadjm-Tehrani. *Crowdroid: Behavior-based malware detection system for Android*. In Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices. Chicago, USA. 2011: 15-26.