A REVIEW ON IDENTIFICATION AND CLASSIFICATION OF MALWARE ATTACKS

Deepa Kolar
Department of Computer Science and Engineering, Basaveshwar Engineering College,
Bagalkot - 587103, Karnataka State, India

Praveen Challagidad

Department of Computer Science and Engineering, Basaveshwar Engineering College,
Bagalkot - 587103, Karnataka State, India

Abstract: Malware variants identification and classification is the one of the most important research problem in digital forensics. Statistic and dynamic analysis is used for detection and classification of malware. Smart phones are being utilized by a vast majority of users for every day planning, data exchanges, correspondences, social interaction, business execution, bank transactions, and almost in each walk of everyday lives. With the expansion of human reliance on smart phone technology, cyber attacks against these devices have surged exponentially. This paper presents a review papers on identification and classification of malware attacks and summary of the research work which have been carried out by many researchers in this field. The through literature review of 10 papers is presented here.

Keywords— Malware family, Malware category, Deep AMD, Machine learning, Cyber attack Security

I.Introduction

The malevolent programming is malware. It is an executed on a registering gadget, alludes to programming programs intended to harm. In a PC framework they do any sort of undesirable activities. Then individuals Cybercriminals attack the systems and organizations with various objectives. They are destroying operating systems, encrypting sensitive data and gathering personal information. To protect users like computers, mobiles, servers and gateways this tools should be developed and deployed to prevent attacks. In cyber security field one of the most crucial problems is malware classification. Then the adware, ransom ware, scare ware, and SMS malware these four are the dataset comprises in malware in this comprises different identification approaches are being proposed.

Scientists have detailed two unique ways to deal with recognize malware. First is Static examination, in which applications are checked without their execution and the second is dynamic investigation, in which malware conduct is dissected in a separated climate after execution. The new assaults are rising each day having the capacity to disregard the security of the Cell phone. The security strategy can be disregarded in various manners as indicated by the working framework being utilized by the Cell phone. Yet, they have certain restrictions by giving an exceptionally effective way to deal with the location and ID of kinds of malware.

Our proposed approach named Deep AMD for noxious application identification. Profound learning is a part of AI that endeavors to gain undeniable level highlights straightforwardly from the first information. Deep AMD includes highlight extraction, order of utilization as vindictive or kind, arrangement of malware class, and malware family. To start with, in the Static layer, the examples are named malware and considerate and investigating to malware applications. Another is Dynamic layer, the examples that are characterized by the Static layer as malware are additionally ordered in to various 4 classes (1) Adware, (2) Payoff product, (3) Scare ware, and (4) SMS Malware.

1) Adware:-

The lone reason for this malware type is showing notices on the PC. Frequently adware can be viewed as a subclass of spyware and it will impossible lead to emotional results. In a few cases, an Adware may hack the Cell phone speaker adware can be used intentionally by a commercial organization, other intruding adware may moreover abuse a notice organization and undermine pay and information from the owners of the advancement organization. Forceful adware can make backup courses of action on to the home screen, take book-marks, change default web program, web crawler, web settings, and pushing trivial notification Microscopic fish is one such sort of forceful adware. Adware can be intended to assume responsibility for the client's android gadget when it is converged with botnet and repacks itself as a mainstream application.

2) Ransom ware:-

This sort of malware intends to scramble all the information on the machine and request that a casualty move some money to get the unscrambling key. Typically, a machine infected by ransom ware is "frozen" as the client can't open any file, and the work area picture is utilized to give data on assailant's demands. Android deliver product regularly fits the overall importance of a diversion. On occasion, the malicious APKs copy simply the name and image of the ordinary application or cover it as a legitimate report in a SMS or email.

3) Scare ware:-

Scare ware is the kind of malware that intends to unnerve the end-client paying for futile applications Scare ware is made to panic or to trap clients with some phishing site to take their data Scare ware stunts the client by introducing trick as genuine applications that regularly assume the presence of safety applications.

4) SMS malware:-

SMS assaults include the creation and conveyance of malware by cybercriminals intended to focus on a casualty's cell phone. SMS malware assumes responsibility for messages on an Android gadget and sends undesirable messages. This malware is being utilized by the proprietor to send messages by the tainted cell phones on his order. SMS assaults incorporate the making and spreading of malware by assailants, planned to zero in on a hacked person's advanced mobile phone utilized by the digital criminal for making enormous earnings groups for cybercriminal frameworks.

Aim: To design and develop a technique for identification and classification of malware attacks using deepAMD .

Advantages in the use of cyber security objects:

1) Protects system against viruses, worms, spyware and other unwanted programs.
2) Protection against data from theft.
3) Protects the computer from being hacked.

II. METHODOLOGY

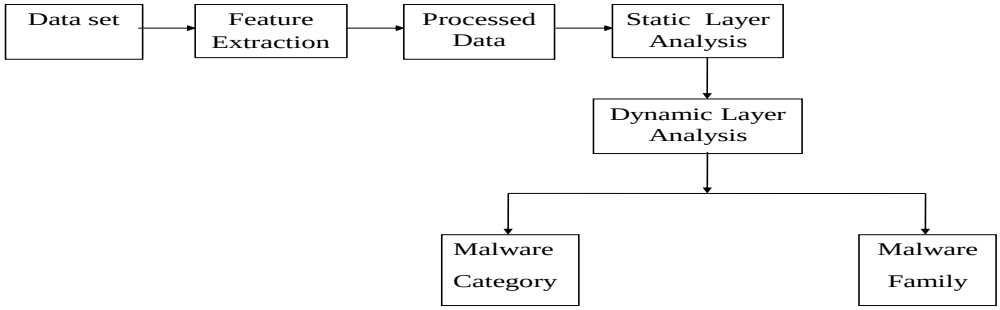

Figure 1

FLOW OF WORK:

CICINVESANDMAL2019 Dataset:

The information is gathered when a dataset is given as info. Now and then the information gathered may contain superfluous information and be put away in a database. The dataset likewise incorporates Malware tests in this dataset are ordered into four classifications: (1) Adware, (2) Ransom ware, (3) Scare ware, and (4) SMS Malware. Information is gone through the pre-preparing stage for additional element extraction.

Pre-processing:

Pre-preparing is a fundamental advance to acquire the best presentation in any AI model. When the information are gathered, regularly the information won't be in a structure that is reasonable for handling. To make the information appropriate for preparing, it is fundamental to change them into an arrangement i.e., the information dataset may not be in the configuration first, we should change over/change in a reasonable organization utilizing information mining calculations, for example, multidimensional, time arrangement, or semi-organized configuration. Pre-handling incorporate expulsion of Na N, and standardization/scaling. The chose dataset has low change and ambiguities along these lines, we pick Min Max scaling for highlight standardization. Normalizing alludes to their scaling of genuine esteemed numeric properties to a fixed reach. It is profoundly imperative to scale the information ascribes for a model that depends on the greatness of qualities. Blend Max scaling standardizes information utilizing the equation referenced in Where

$$X\ norm = \frac{Xi - X\ min}{X\ max - X\ min}$$

Xi is the original value of the feature that is subtracted from minimum value of that feature and divided by subtracted result of max and min of the feature.

Feature extraction:-

Highlight extraction used to CICInvesAndMal2019 dataset is a two-layered dataset comprising of Static layers and Dynamic layers for malware discovery. One layer is the Static layer used to investigate malware applications and assuming it identifies any noxious application, that application is likewise considered malware for the Powerful layer. In the event that the Static-base first layer identifies a dubious malware, there is greater chance of pernicious goals around there. We accept if an example is distinguished dubious with a Static layer, the analyzer ought to consider it as malware for the following layer also. Subsequently, we can decrease the danger of confiding in obscure examples. In the dataset, there are trying and preparing tests of both Static layers and Dynamic layers.

Machine learning models:-

we use the following machine learning algorithms to evaluate and compare the effectiveness so four proposed Deep AMD approach:-

1) Naive Bayes (NB)
2) Sequential Minimal Optimization
3) Multilayer Perceptron
4) Decision tree
5) Deep Artificial Neural Networks

This algorithm works to classify of malware category, and malware family.

III. LITERATURE REVIEW

Malware Location. Long Wen [1] this paper gives to AI based light weight framework that is equipped for recognizing malware on Android gadgets. In this framework separate highlights dependent on the static examination and the dynamitic investigation, at that point another element choice methodology dependent on guideline segment investigation (PCA) and alleviation are introduced in the article to diminish the elements of the highlights. From that point forward, a model will be built with help vector machine (SVM) for arrangement. Test results show that our framework gives a powerful strategy in Android malware discovery. And afterward another element determination calculation named PCA-Help is proposed to track down the most segregating highlight subset. Test results show that our technique can precisely recognize Android malware. AbdelmonimNaway [2] this paper gives to profound learning has been acquainted in with malware identification. Profound learning is a part of AI that relies upon considering different degrees of portrayals, practically equivalent to a positioning of highlights or thoughts, where high level ideas are resolved from lower-level ones, and comparable lower-level ideas could help with deciding various high level ideas. Profound learning for malware order offers a methods for building versatile AI models, which may deal with any proportion of information, without exhausting reliably of assets like memory. Profound learning marks malware rely upon the overall example, which coordinates the recognizing of an assortment of malware assaults and their varieties. Besides, profound learning conducts a significant grouping and improves its exactness since profound learning recognizes a larger number of highlights than regular AI techniques by going through numerous degrees of highlights extraction. This empowers profound learning models to procure another example of malware after the fundamental preparing stage. Zhejiang Wang [3]This paper gives to Android working framework involves the main piece of the pie and its open-source nature, its security is progressively compromised and tested. To manage the touchy development of Android malware, research on Android malware identification utilizing profound learning has been an exploration area of interest as of late, however right now deficient with regards to an itemized and exhaustive prologue to the examination progress of Android malware location utilizing profound learning. Right off the bat, the paper presents the fundamental information on Android malware location innovation. At that point it sort outs, breaks down, and sums up Android malware discovery's most recent examination progress dependent on profound learning and sums up the Android malware identification design and location conspire dependent on profound learning. At that point it examines the issues and difficulties of Android malware location utilizing profound learning. MattiJuutilainen [4] this paper gives to malware recognition is a significant factor in the security of the PC frameworks. Nonetheless, at present used mark based techniques can't give precise recognition of zero-day assaults and polymorphic infections. To decide the best component extraction, highlight portrayal, and arrangement strategies that outcome in the best exactness when utilized on the highest point of Cuckoo Sandbox. In particular, k-Closest Neighbors, Choice Trees, Backing Vector Machines, Innocent Bayes and Irregular Backwoods classifiers were assessed .The dataset utilized for this investigation comprised of the 1156 malware records of 9 groups of various kinds and 984 considerate documents of different configurations. This work presents suggested strategies for AI based malware grouping and location .This paper examines the central matters and worries of AI based malware discovery, just as searches for the best component portrayal and characterization techniques. Maria J. Schroeder[5] This paper gives to Present day antivirus programming is successful at distinguishing realized dangers yet can be sidestepped by uncommonly created novel malware, for example, polymorphic malware, which can reinvent itself to show up and work contrastingly while eventually playing out similar capacities generally. Conventional antivirus programs use signature-based identification, which includes checking potential malware against a data set of hashes fingerprints of the specific documents of known malware, yet this strategy has serious limits .Current antivirus items likewise utilize both static heuristic checks and dynamic investigation, which utilize physically created rules to identify malware dependent on code design or conduct. Nonetheless, an assortment of methods have been distributed that rout both heuristic and dynamic examination. Azizalotaibi [6] This paper gives to advanced cell applications dependent on the Android operating system stage is quickly developing among PDA clients. , vindictive applications for Android are being created to perform assaults, for example, annihilating working frameworks, taking private information, gathering individual data, and commandeering or scrambling touchy information. A few malware identification frameworks dependent on AI have been created and sent to extricate an assortment of highlights to forestall such assaults This paper proposes a novel structure, in particular, MalRes LSTM, in light of profound lingering long momentary memory to distinguish and arrange malware variations. The system forces a bunch of requirements on the profound

learning engineering to catch conditions between the removed highlights from the Android bundle unit (APK) document. These capabilities are planned to a vector space to deal with the information grouping utilizing a succession model dependent on the leftover LSTM organization. To assess the exhibition of the proposed structure, a few investigations are directed on the Drebin dataset .Cristian Pascariu[7] this paper gives to record how to apply AI calculations to the dynamic malware examination measure. This is vital, as tainted documents have conditions on different cycles that are running on a framework. The objective is to recognize explicit models where malware will use safe projects and utilities to introduce itself, these models of vindictive exercises will be utilized to prepare a neural organization, so that other PC infections that utilization a similar method would be distinguished. This arrangement is centered around recognizing malignant infections on PCs that run Windows as a working framework, this is a famous and has a high introduce rate inside the customer market just as people in general and private areas. This paper will zero in on recognizing examples of malevolent conduct and safe conduct that will be utilized to prepare a fake neural organization to identify PC infections and dispense with bogus positives. Liminshen[8] This paper gives to shrewd activity framework, Android is the most mainstream utilized portable stage for advanced cells and IoT .static malware recognition dependent on authorizations, purpose and part data. The static highlights datasets are contribution to a completely associated neural organization to recognize the malware and test its adequacy through tests. Organization traffic based powerful versatile malware recognition. Our trial results show that consolidating network traffic highlights with falling profound learning CACNN techniques can viably recognize malevolent programming in Android Applications. Two-layer discovery model. The principal layer, applying a completely associated neural organization to dissect static highlights, and info the outcomes to the following identification layer. Second layer, network traffic highlights recognition broke down the eventual outcomes to demonstrate that CACNN model can successfully recognize malware. Simultaneously, this models can likewise recognize malware by its class and malignant family. In general, consolidating two layer of discovery model can additionally improve the recognition proficiency. Deepika P Vinchurkar[9]This paper gives to Pernicious clients and saltines look for powerless targets, for example, un patched frameworks, frameworks contaminated with Trojans, and organizations running unreliable administrations The affirmation of trustworthiness and wellbeing ought to be applied to PC frameworks and information. The Web has made the data stream to the enormous degree. Likewise simultaneously it needs to confront numerous dangers and assaults. Subsequently the security alert is needed to control the assaults and dangers. A warning should be shipped off the directors and security colleagues about the different dangers and assaults which has happened so they can react progressively to the danger. The paper depicts the difficulties in IDS. In this paper we talk about different strategies for inconsistency recognition procedures and spotlight on the AI based methods. Guideline Part Investigation calculation to decrease dimensionality is depicted in the paper. The paper likewise centers around characterization utilizing Backing Vector Machines. Divya Bansal [10] this paper gives to Programming that "purposely satisfies the unsafe goal of an aggressor" is alluded to as malevolent programming or malware. To beat the constraint of mark based strategies, malware investigation methods are being followed, which can be either static or dynamic. The malware investigation methods assist the experts with understanding the dangers and intensions related with a vindictive code test. The insight so obtained are often wont to react to new trends in malware development or take preventive measures to deal with the threats coming in future.. Features derived from analysis of malware can be used to group unknown malwares and classify them into their existing families. This paper presents a review of techniques approaches for analyzing and classifying the malware executables. This paper highlights the prevailing techniques for analyzing, detecting and classifying malwares. The list of publications categorized consistent with malware analysis techniques. The machine learning technologies that are getting used in detecting and classifying malwares aren't capable handle challenges arising from the large amount of dynamic and severely imbalanced network data. These should be transformed so that their potential can be leveraged to address the challenges posed in cyber security.

Table I Literature Survey Table

Author	Description	What they used	Drawback
Long Wen	Focuses on a classification model by using SVM and evaluate the unknown Android application by classifying into malware or benign.	PCA-RELIEF algorithm easy to classify the malwares.	
.AbdelmonimNaway	To organize an inclusive review of the work accomplished in Android malware analysis using deep AMD with static analysis, dynamic analysis and hybrid analysis	deep convolution neural network are used	It consumes the Android system resources and takes a long time to perform the analysis.
Zhejiang Wang	The method's performance metric is better than the method based on traditional machine learning technology.	A method based on support vector machine ,random forest, logistic regression, and K-nearest neighbor algorithm are used	
Matti Juutilainen	the machine learning based malware classification based on Cuckoo Sand box. This sandbox is going to be utilized for the extraction of the behavior of the malware samples, which can be used as an input to the machine learning algorithms.	K-Nearest-Neighbors, Decision Trees, Support Vector Machines, Naive Bayes and Random Forest classifiers were evaluated	Its main drawback is that every feature is treated independently, although in most cases this can't be true.
Maria J. Schroeder	These works explore the drastic performance drop which occurs when using a realistically low proportion of malware in test datasets instead of datasets split evenly between malware and benign software.	Naïve Bayes, Logistic Regression, Classification and Regression Tree, Rando Forest algorithms are used.	the advantage of a class must be demonstrated over a range of training and test malware prevalence's.
Aziz alotaibi	This paper proposes MalResLSTM, based on deep residual long short-term memory to identify and classify malware variants.	Convolution neural network (CNN) and recurrent neural network (RNN). Algorithms are used	

Cristian Pascariu	Identifying patterns of malicious behavior and safe behavior that will be used to train an artificial neural network to detect computer viruses and eliminate false positives.	Artificial neural networks are used.	Detecting malicious viruses on computers that run Windows as an operating system.
Limin shen	Two layer model for detecting malware. CACNN model is used to classify malware from the benign. CACNN model can also classify Malware by category and family.	Convolution Auto-Encoder algorithms are used.	Mobile APPs make our life more convenient, it also brings enormous burden to the mobile and IoT security protection.
Deepika Vinchurkar	Intrusion detection system can work by observing the unauthenticated and unauthorized use of different programs of networking.	Genetic algorithm, SVM algorithm are used	
Divya Bansal	The provides an overview of techniques for analyzing and classifying the malwares.	Clustering algorithm are used	

IV. ISSUES AND CHALLENGES

ISSUES

- Malware are being used to attack critical infrastructures, for espionage against a nation, for stealing private information or for conducting financial frauds. The malware analysis and detection recently usedby many researchers for visualization technique to understand malware visually that can help antivirus software to detect malware effectively.

CHALLENGES

- The main objective is to identify and detect the types of malwares.
- To perform malware categorization using Neural network and Deep AMD .
- To perform dynamic analysis using classify the malware's category and malware 's family.

V. CONCLUSION

deepAMD, a successful orderly and useful way to deal with distinguish and recognize Android malware, malware class, and family on both static and dynamic layers. Android malware is being recognized and malware is developing. To counter malware in Android gadgets, in the Static layer to arrange twofold malware and afterward classify malware utilizing DeepAMD, for malware family characterization on the Unique layer. The DeepAMD is assessed utilizing the cutting edge CICAndMal2019 dataset and exploratory outcomes showed that DeepAMD is the most proficient technique for recognizing and distinguishing malware on the Static just as Powerful layer. This paper centers around tending to every one of these constraints by giving a profoundly productive way to deal with the location and ID of kinds of malware.

REFERENCES

1) Long Wen," An Android malware detection system based on machine learning" Beijing University of Technology, Beijing 100049, China.[2017]

2) Abdelmonim Naway "A Review on the Use of DeepAMD in Android Malware Detection" 2 Beinong Road, Champing District, Beijing, China, 102206.[2016]

3) Zhejiang Wang" Review of Android Malware Detection Based on DeepAMD" Technology Institute, Beijing 100071, China2State Information Center, Beijing 100000, China.[2020]

4) Kateryna Chumachenko" Machine learning methods for malware detection and classification" [2017]

5) Midshipman 1/C Zane" Machine learning based malware detection" [2015]

6) Aziz alotaibi "Identifying Malicious Software Using Deep Residual Long-Short Term Memory" Taif University, Taif 21974, Saudi Arabia.[2019]

7) Cristian Pascariu "Dynamic analysis of malware using artificial neural networks" UPB Bucuresti, Romania.[2017]

8) Jiayin Feng "A Two-Layer DeepAMD Method for Android Malware Detection Using Network Traffic" Yan Shan University, Qinhuangdao 066004, China.[2020]

9) Deepika P Vinchurkar, "A Review of Intrusion Detection System Using Neural Network and Machine Learning Technique" MPSTME, SVKM's NMIMS University Mumbai, India.[2012]

10) Divvy Bansal,"Malware Analysis and Classification" PEC University of Technology, Chandigarh, India.[2014]