iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VIII **Month of publication:** August 2017

DOI: http://doi.org/10.22214/ijraset.2017.8294

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

Detection and Classification of Malware Data in Android Based Smart Device

Sandeep Sharma[1], Ruchi Jain[2], Babita Pathik[3], Dr. Shiv K. Sahu[4]
Dept. of CS & Engineering LNCTE, Bhopal (M.P.)

Abstract: ***The detection and classification of malware data in android based smart device is very serious challenge due to self-propagation nature of malware. Most of the malware nature is dynamic and self-propagated and infected automatically files and consume more energy and switch off the smart devices. The infected files also decrease the performance of smart devices. In this paper proposed novel methods for the detection and classification of malware data in android based smart devices. The proposed algorithm is combination of support vector machine and wrapper feature extraction of malware data. the wrapper feature extractor extracts the malware features in three level, application, network and processing of file. The proposed algorithm implemented in android visual studio and used SSMD malware dataset. The proposed algorithm is better in comparison of SVM.***
Keywords: Smart devices, malware, android, classification, detection.

I. INTRODUCTION

Malware classification and identification process is extremely unpredictable process in smart device. In current survey device situation, different sorts of malware family are accessible some are known family and some are obscure family. The group of know malware identification utilized some understand strategy, for example, signature based method and control based procedure [1, 2]. If there should arise an occurrence of obscure malware group of assault identification is different testing undertaking. In current pattern of malware identification utilized a few information mining methods, for example, grouping and bunching. The procedure of grouping enhances the procedure of discovery of malware. In this exposition utilized diagram based system for malware characterization and discovery. The coherence of section talks about element extraction procedure of malware information, coordinated non-cyclic diagram strategy, bolster vector machine and proposed technique. Malware grouping and recognition process is extremely intricate process in brilliant device. In current keen device situation, different sorts of malware family are accessible some are known family and some are obscure family [3, 4]. The group of know malware location utilized some surely understand system, for example, signature based strategy and control based method. in the event of obscure malware group of assault recognition is different testing assignment. In current pattern of malware discovery utilized a few information mining strategies, for example, grouping and bunching. The procedure of grouping enhances the procedure of location of malware. Ruining malware assaults in keen devices is a flourishing exploration territory with a generous measure of still unsolved issues. Because advanced mobile phones, one essential line of protection is given by the security engineering of the device, one of whose chief elements is an authorization framework that confines applications benefits. This has demonstrated patently deficient in this way. For instance, on account of ANDROID OS applications ask for authorizations in a non-debatable form; in a manner that clients are left with the decision of either conceding the application all that it requests at establishment time or it won't be conceivable to utilize it. Most clients just don't focus on such demands; or don't completely comprehend what every authorization implies; or, regardless of the possibility that they do, it is difficult to make sense of every conceivable outcome of allowing a given arrangement of benefits [7, 8]. The rest of paper discuss as section feature extraction of malware data. in section III. Discuss the proposed method of malware classification. In section IV experiment result analysis and finally discuss conclusion and future scope.

II. FEATURE EXTRACTION

Malware order can either have single variable approach or a multi-variable way to deal with recognize Malware relying upon the calculation utilized. In the single variable approach a solitary variable of the framework is broke down. This can be, for instance, port number, CPU use of a nearby machine and so on. In multi-variable approach a blend of a few components and their between relationships are examined. likewise in view of the strategy the route in which components are decided for the IDS can be isolated into two gatherings; into highlight choice and highlight diminishment. In the component choice technique the elements are either picked physically from the information checked or by utilizing a particular element determination device [10, 11]. The most reasonable components are chosen by handpicking from the element range in light of the earlier information about the environment that the IDS are observing. For instance, includes that can recognize certain sort of activity from the movement streams are picked

for the system activity display preparing. The thought behind the component choice instruments is to diminish the measure of elements into a practical subset of elements that don't associate with each other. Cases of highlight choice instruments are Bayesian systems (BN) and characterization and relapse tree (CART). Bayesian system is a probabilistic graphical model that speaks to the probabilistic connections between components. Truck is a strategy that utilizations tree-building calculations to develop a tree-like if-then expectation designs that can be utilized to decide distinctive classes from the dataset. [12].

Categories and Numbers of Extracted Features [14]

Source	Category	#Feature
Dynamic	File operations	19038
Static and dynamic	Signatures	1283
Dynamic	Registered_receivers	8334
Dynamic	Reflection_calls	14799
Static	Used/required permissions	1267
Static and dynamic	SMS, phone, contacts	1493
Static	Application components	21523
Static and dynamic	Dynamic code loading	916
Static and dynamic	Crypto operation	41
Dynamic	Data_leak	828
Dynamic	Commands	937
Static and dynamic	Network activity	37734
Static	The use of special API	20162
Dynamic	System properties	13081

III. PROPOSED METHODOLOGY

The proposed algorithm of malware classification is combination of classification and feature selection process. For the selection of feature used wrapper algorithm and process of classification used support vector machine. the support vector machine used radial kernel function for the groping of selected features. the processing of classification technique is describing here.

A. Input the selected feature from wrapper in support vector machine

B. The selected value of features F mapped in relation of Fi$\in R^d$ here d is size of feature vector

C. Sampling of input features vector for the measuring similarity of features sim features$= \sum_{i=1}^{m} \sum_{j=1}^{n} sim(Xi, xj)/m * k$

D. Derive the kernel function of real valued function K(x,y)

E. Measure the value of hypothesis of kernel function

$x(t) = w0 + \sum_{j=1}^{total\ data} wj\ exp(\frac{-(total-xj)}{\sigma^2})$ this kernel function of SVM

F. Used product function of similar data in (Fx.FY) and estimate final C data for training of class

G. Lo=SimFet$\frac{1}{p}\sum_{i=1}^{p} min\ (FX - Fy)$ where Lo is training parameter of kernel function.

H. Assigned the trained feature as label c1, c2,……cn If class level is C=¢ then terminated the process there is no features for the classification

I. Else Find level of class in terms of product of similarity

C=R*X^d

J. Estimate margin value of kernel for similar class

K. End the process and feature are classified

L. Malware is detected

IV. EXPERIMENTAL RESULT

For the assessment of proposed model utilized java advancement programming, java include, java apparatuses and android improvement unit for the preparing of malware characterization. For the malware characterization, we utilize a few datasets as CSDMC2010_API.tar.bz2 and readme_Task3.txt. The relative database downloads from www.csmining.org. The proposed calculation is adjustment of Machine Learning [15].

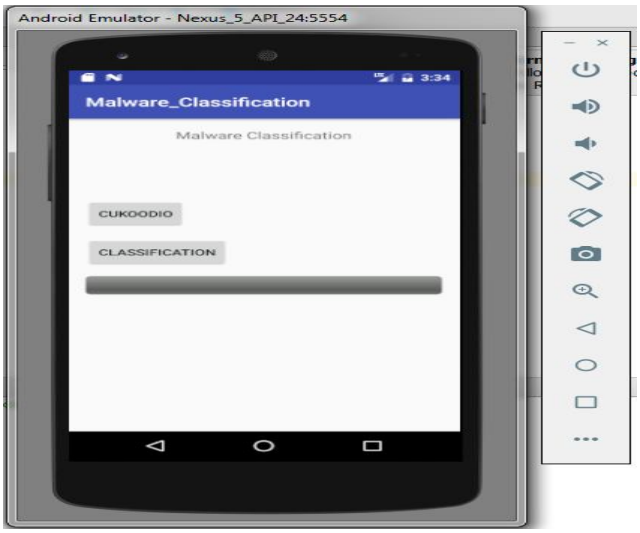


Figure 1: show that the window of malware classification open screen with classification button and process bar in our malware classification implementation.

Figure 2: show that the window of proceed process bar in malware classification screen with classification button and process bar processed in our malware classification implementation.

SUPPORT VECTOR MACHINE			
Categories Of Malicious Dataset	Detection Rate	Precision Rate	Recall Rate
Crypto operation	89.79	81.93	80.93
Data leak	88.79	80.83	78.46
Commands	86.79	79.56	81.93
Network activity	85.79	82.93	81.93
The use of special API	86.79	84.43	79.93

Table 1: Shows that the performance evaluation of Detection rate, Precision rate and Recall rate for Support Vector Machine Learning method

PROPOSED LEARNING METHOD			
Categories Of Malicious Dataset	Detection Rate	Precision Rate	Recall Rate
Crypto operation	95.80	85.02	83.97
Data_leak	93.83	81.97	80.97
Commands	94.83	84.97	81.97
Network activity	95.67	85.97	84.97
The use of special API	92.83	86.94	82.94

Table 2: Shows that the performance evaluation of Detection rate, Precision rate and Recall rate for proposed method.

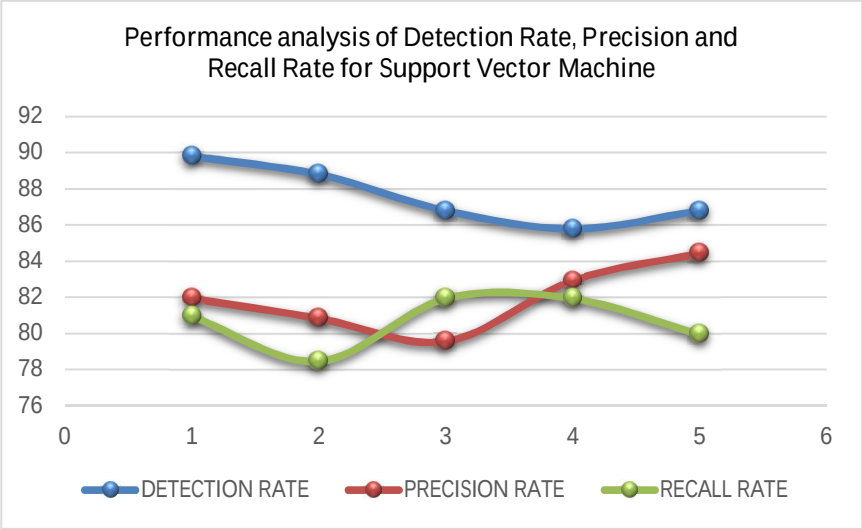


Figure 3: Shows that the performance evaluation of Detection rate, Precision rate and Recall rate for Support Vector Machine.

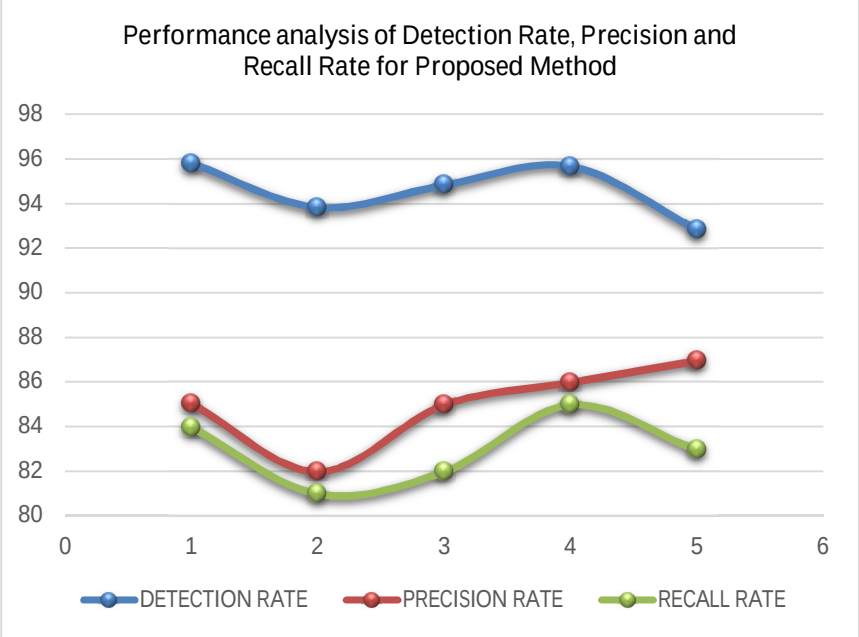


Figure 4: Shows that the performance evaluation of Detection rate, Precision rate and Recall rate for proposed method.

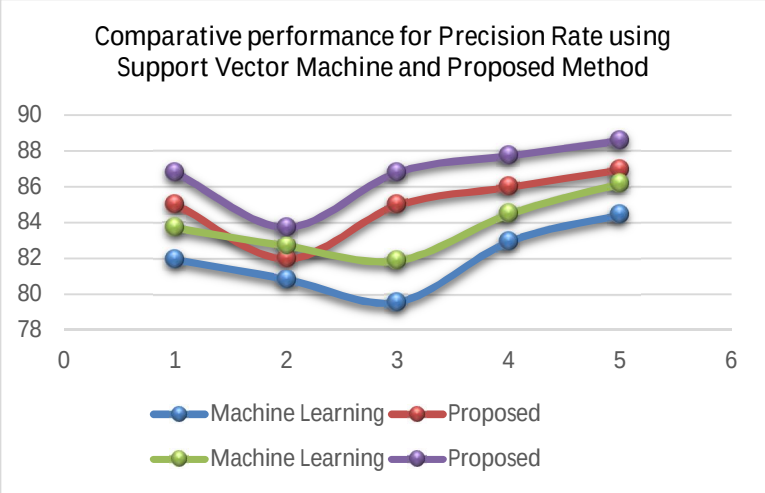


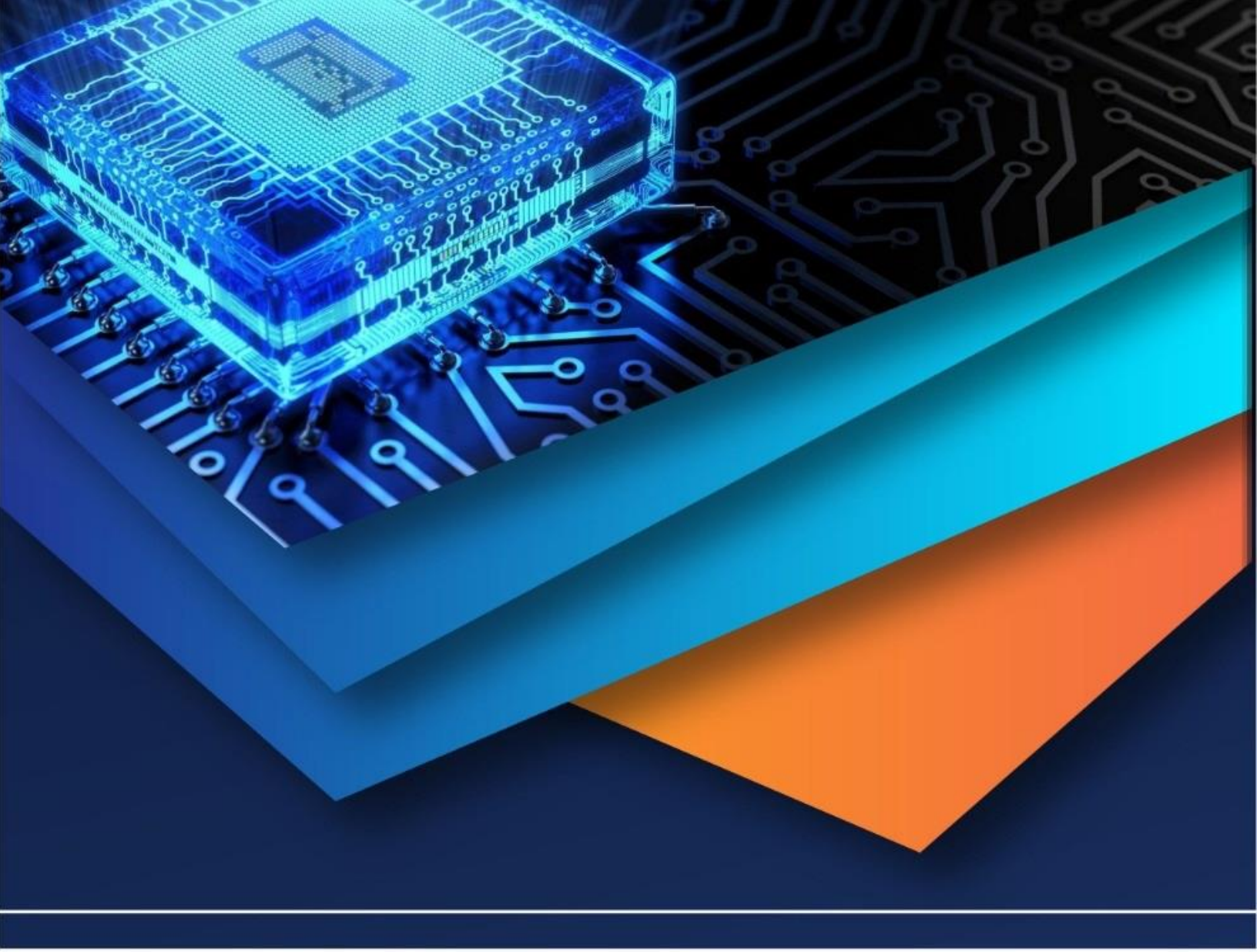
Figure 5: Shows that the performance evaluation of Precision rate for Support Vector Machine and Proposed method.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel algorithm of malware classification, in view of wrapper and Gaussian Support Vector Machines, for malware order. Explores different avenues regarding the malware information family Data demonstrate that SVM-DAG can give great speculation capacity and adequately grouped malware information. In addition, the adjusted calculations proposed in this despoiling beat ordinary machine learning and cuckoo based calculation as far as accuracy and review. In particular, precision of the changed calculations can be increment because of highlight portion of feature extractor, and decreases include sub set increment the exactness of order. From our tests, the SVM can distinguish known assault sorts with high exactness and low false positive rate which is under 1%. The proposed strategy grouped assault and typical information of malware family information is precisely. In the accumulation of highlight quality of malware information, some suspicious components are not gathered, so in future utilized heuristic capacity for better determination of elements.

REFERENCES

[1] Ke Xu, Yingjiu Li and Robert H. Deng "ICCDetector: ICC-Based Malware Detection on Android", IEEE, 2016, Pp 1252-1264.
[2] Guillermo Suarez-Tangil, Juan E. Tapiador, Flavio Lombardi and Roberto Di Pietro "ALTERDROID: Differential Fault Analysis of Obfuscated Smartphone Malware", IEEE, 2016, Pp 789-802.
[3] Luca Caviglione, Mauro Gaggero, Jean-François Lalande, Wojciech Mazurczyk and Marcin Urbanski "Seeing the Unseen: Revealing Mobile Malware Hidden Communications via Energy Consumption and Artificial Intelligence", IEEE, 2016, Pp 799-810.
[4] Lilian D. Coronado-De-Alba, Abraham Rodríguez-Mota and Ponciano J. Escamilla- Ambrosio "Feature Selection and Ensemble of Classifiers for Android Malware Detection", IEEE, 2016, Pp 1-6.
[5] Yu Feng, Saswat Anand, Isil Dillig and Alex Aiken "Apposcopy: Semantics-Based Detection of Android Malware through Static Analysis", ACM, 2014, Pp 1-12.
[6] Prof. Amruta Gadekar, Sharad Goykar, Shesharao Chatse and Vishaka Deore "A Survey on a ICC-Based Malware Detection on Android", IJETCS, 2016, Pp 1-5.
[7] Suleiman Y. Yerima, Sakir Sezer, Gavin McWilliams and Igor Muttik "A New Android Malware Detection Approach Using Bayesian Classification", IEEE, 2013, Pp 1-8.
[8] Mingshen Sun, Xiaolei Li, John C.S. Lui, Richard T.B. Ma and Zhenkai Liang "Monet: A User-oriented Behavior-based Malware Variants Detection System for Android", arXiv, 2016, Pp 1-13.
[9] Hugo Gascon, Fabian Yamaguchi, Daniel Arp and Konrad Rieck "Structural Detection of Android Malware using Embedded Call Graphs", ACM, 2013, Pp 1-10.
[10] Borja Sanz, Igor Santos, Carlos Laorden, Xabier Ugarte-Pedrero, Javier Nieves, P
[11] ablo G.Bringas and Gonzalo Álvarez "MAMA: Manifest Analysis for Malware Detection in Android", ACM, 2013, Pp 1-19. Pengbin Feng, Jianfeng Ma and Cong Sun "Selecting Critical Data Flows in Android Applications for Abnormal Behavior Detection", Springer, 2017, Pp 1-14.
[12] Yousra Aafer, Wenliang Du and Heng Yin "DroidAPIMiner: Mining API-Level Features for Robust Malware Detection in Android", Springer, 2014, Pp 1-18.
[13] Borja Sanz, Igor Santos, Carlos Laorden, Xabier Ugarte-Pedrero, Pablo Garcia Bringas and Gonzalo ´Alvarez "PUMA: Permission Usage to detect Malware in Android", Springer, 2013, Pp 1-10.
[14] Nirmala Yadav,Aditi Sharma and Amit Doegar "A Survey on Android Malware Detection", IJNTR, 2016, Pp 47-53.
[15] Geethu M Purushothaman, G Gopinadh and Nihar SNG Sreepada "Malware Detection in Android", IJARCET, 2014, Pp 1429-1436.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY