A unique technique for signature Based Malware Detection UsingMachine Learning Techniques

[1]Syed Amreen, [2]Y Laxmikanth, [3]P Swaroopa, [4]Jennepally Sarika Reddy
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer Science Engineering, Visvesvaraya College of Engineering & Technology, Hyderabad, India.

ABSTRACT
The quantity of malwares is increasing incredibly quickly, and because of this, computer security researchers are forced to develop new methods of securing networks and PCs. One of the most popular methods for defending against software attacks aimed at your computer is signature-based detection. Viruses, malware, worms, Trojan horses, and other hazards are among them. Additionally, there are two types of malware analysis—static and dynamic—that are typically used to detect malicious software. Malicious software, malicious code (MC), and Malcode are terms used to describe software that crashes or disrupts regular operations without the user's awareness. Antivirus software uses a database together with signature-based detection. They will look for computer scan results that match known malware traces. The traces of this trojan are kept in a database. This type of detection involves your antivirus having a predefined repository of static signatures that represent known network threats. These threats are different from one another because of their unique coding. Any malware signature that matches the database will be detected on the system.
Keywords: Signature Based Detection, Malware

I. Introduction

Malicious software, malicious code (MC), and Malcode are all terms for software that crashes or breaks regular operations without the user's awareness.
The quantity of malwares is increasing incredibly quickly, and because of this, computer security researchers are forced to develop new methods of securing networks and PCs. One of the most popular methods for dealing with computer software risks is signature-based detection. These dangers include Trojan horses, worms, Trojan horses, and viruses. Computers need to be shielded against an enormous number of threats.
Simply by detecting the signature of any dangerous file contained in the database, signature-based antivirus, as a type of malware detection approach, has the capacity to find and eliminate any known malware. Achieving this protection is hugely dependent on a well-crafted, advanced, signature- based detection being at the helm of affairs.

II. SYSTEM DESIGN

UML, short for Unified Modeling Language, is a standardized modeling language consisting of an integrated set of diagrams, developed to help system and software developers for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing object-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects. Using the UML helps project teams communicate, explore potential designs, and validate the architectural design of the software. In this article, we will give you detailed ideas about what is UML, the history of UML and a description of each UML diagram type, along with UML examples.
GOALS: The following are the primary design goals of UML: A consistent, user-friendly, descriptive language that people can use to build models and share them. Provide mechanisms to extend and special ize the core concepts.Operate freely regardless of the language or process. This formal modelling language understanding has a basis in how it is structured. Boost the development of OO toolmakers.
System architecture is the structure of an IT system. The architecture of complex systems such as an organization is most typically referred to as business architecture or enterprise architecture. System architecture defines the structure of a software system.

System Architecture

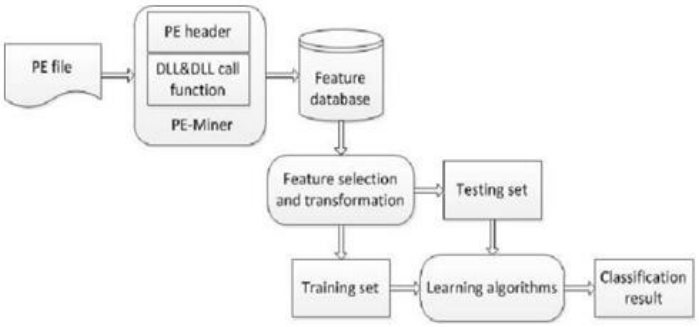

Fig.1.System Architecture

The architecture of the Malware Detection Technique usedto determine whether a system/file has malware or not.

Fig.2. Home page of the Software

Within this screen, we can see that we have a homepage, and we have to login in for the further process

Fig.3. within this screen, we can the user login

In above screen showing the details of the user for the login process Once the details are entered by the user ,wehave to click LOGIN button. We can also create a new account

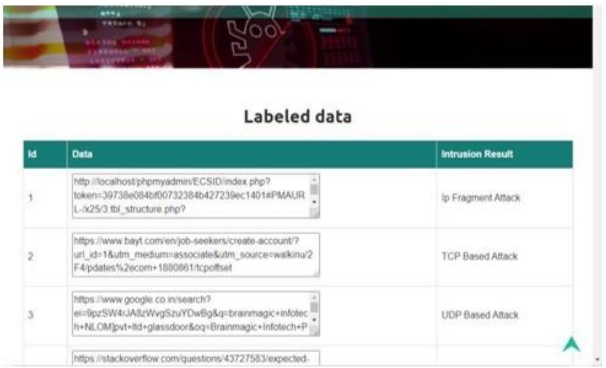

Fig.. 5. In above screen showing the Labeled Data

Within the screen, we have a labeled data and also we have URL's of that labeled data. From the list of that labeled data select one data set and copy that URL.

Fig.6. In above screen selecting and uploading the URL ofDataset.

After selecting the URL of the labeled data upload that dataset url in the add data column and next clicksubmit to analyze the data.

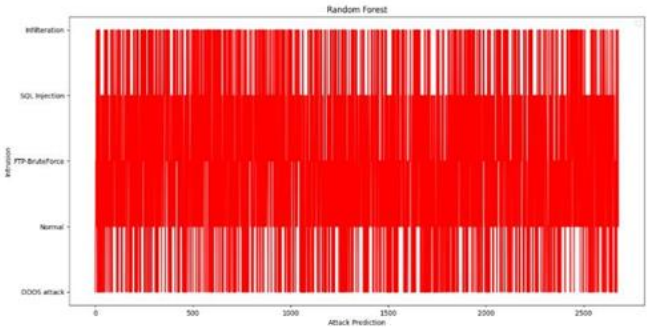

Fig.7. Within the screen showing the graphicalrepresentation of Random Forest

After adding the data set we analyze the data by using the Random Forest Algorithm. The above screen is showing the graphical representation of the data by using the algorithm within the screen showing the accuracy of random forest. The random forest accuracy results display the Train dataaccuracy and Test data accuracy.

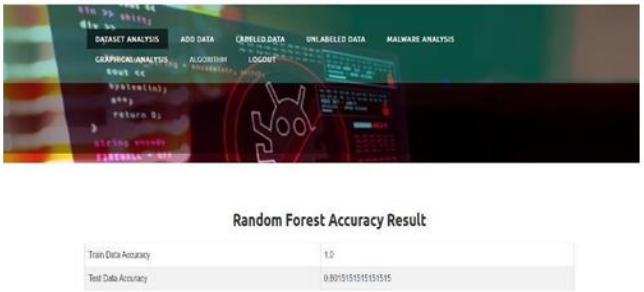

Fig.8 above screen showing the accuracy result of therandom forest.

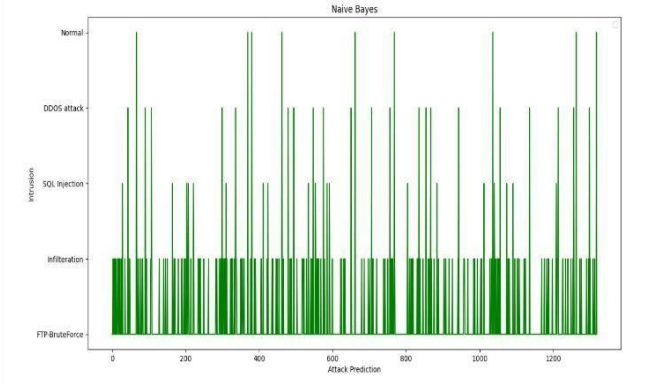

Fig.9. Naïve Bayes Graphical Representation

The above screen is showing the graphical representation ofthe data set by using the Naïve bayes algorithm.

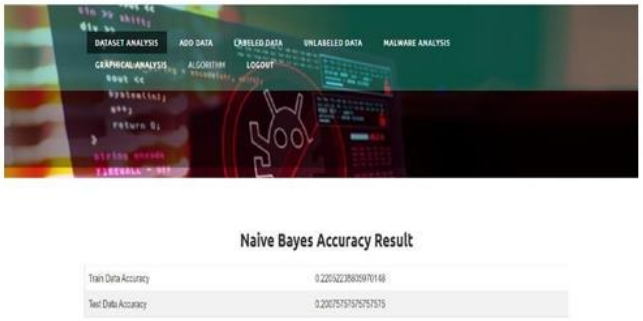

Fig.10 above screen showing the accuracy result of theNaïve Bayes.

Within the screen showing the accuracy of Naïve Bayes . The Naïve Bayes accuracy results display the Train data accuracy and Test data accuracy.

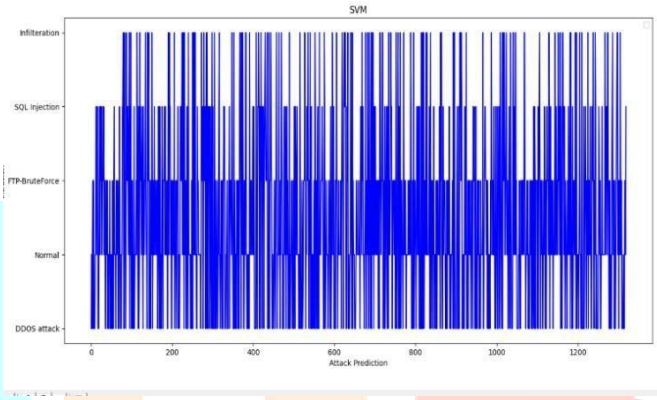

Fig.11. SVM Graphical Representation

The above screen is showing the graphical representation ofthe data set by using the SVM algorithm within the screen showing the accuracy of SVM. The SVM accuracy results display the Train data accuracyand Test data accuracy.

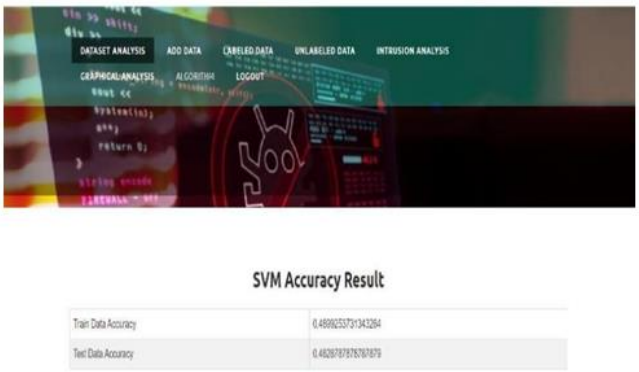

Fig.12 above screen showing the accuracy result of theSVM.

CONCLUSION

Malware detection is viewed as a challenge of classification, where each record may be categorized as either normal or as a specific type of malware. In recent years, machine learning-based malware detection has become increasingly popular. An accurate malware detection model is constructed by selecting an efficient classification strategy as a crucial machine learning application. According to the observed results, the Random forest classifier outperforms other classifiers for the under consideration data-set.

REFERENCES

[1] Zhang, Y., et al., A survey of cyber crimes. Security andCommunication Networks, 2012. 5(4): p. 422-437.

[2] Bazrafshan, Z., et al. A survey on heuristic malware detection techniques. in The 5th Conference on Information and Knowledge Technology. 2013.

[3] La Polla, M., F. Martinelli, and D. Sgandurra, A Survey on Security for Mobile Devices. IEEE Communications Surveys & Tutorials, 2013. 15(1): p. 446-471.

[4] Meng, G., et al., Mystique: Evolving Android Malware for Auditing Anti-Malware Tools, in Proceedings of the 11thACM on Asia Conference on Computer and CommunicationsSecurity. 2016, ACM: Xi'an, China. p. 365-376.

[5] Vemparala, S., et al., Malware Detection Using Dynamic Birthmarks, in Proceedings of the 2016 ACM on InternationalWorkshop on Security and Privacy Analytics. 2016, ACM: New Orleans, Louisiana, USA. p. 41-46.

[6] Dang-Pham, D. and S. Pittayachawan, comparingintention to avoid malware across contexts in a BYOD- enabled Australian university: A Protection Motivation Theory approach. Computers & Security, 2015. 48: p. 281- 297.

[7] Meng, G., et al., Semantic modelling of Android malware for effective malware comprehension, detection, and classification, in Proceedings of the 25th International Symposium on Software Testing and Analysis. 2016, ACM: Saarbrucken, Germany. p. 306-317.

[8] Han, K., J.H. Lim, and E.G. Im, Malware analysis methodusing visualization of binary files, in Proceedings of the 2013Research in Adaptive and Convergent Systems. 2013, ACM:Montreal, Quebec, Canada. p. 317-321.

[9] Grégio, A.R.A. and R.D.C. Santos. Visualization techniques for malware behavior analysis. in SPIE Defense, Security, and Sensing. 2011. SPIE.

[10] Kitchenham, B. and S. Charters, Guidelines for performing systematic literature reviews in software engineering, Technical Report EBSE-2007-01 Ver. 2.3, School of Computer Science and Mathematics, Keele University

[11] K.K., P., B. N.M.W.M., and D.V. N.K., Systematic review: School health promotion interventions targeting physical activity and nutrition can improve academic performance in primary- and middle school children. Health Education, 2013. 113(5): p. 372-391.

[12] Shea, B.J., et al., Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology, 2007. 7(1): p. 10.

[13] Feizollah, A., et al., A review on feature selection in mobile malware detection. Digital Investigation, 2015. 13: p. 22-37.

[14] Ye, Y., et al., A Survey on Malware Detection Using Data Mining Techniques. ACM Comput. Surv., 2017. 50(3):p. 1-40.

[15] Jacob, G., H. Debar, and E. Filiol, Behavioral detectionof malware: from a survey towards an established taxonomy. Journal in Computer Virology, 2008. 4(3): p. 251-266. Programming Interface Call Graph. Vol. 9. 2012. 283-288.