Classification Of Malware Detection Using Machine Learning Algorithms: A Survey

P HarshaLatha, R Mohanasundaram

Abstract: Malware is the one which frequently growing day by day and becomes major threats to the Internet Security. The are several methods for classifying of new malware from the existing signatures or code. The traditional approaches are not much effective to compete the new arriving malware samples. More antivirus softwares provides defense mechanism against malwares but still zero-day attack is not achieved. To enhance in mechanisms machine learning algorithms are used and provide good experimental results accordingly. While the traditional signature approaches are also failed to compete the new malwares. In this paper, we define malware and types of malware as an overview, as well we define the new mechanism of using machine learning algorithms how effective and efficient in classification of malware detection and we presented the existing works related to malware detection classification using machine learning algorithms and it is discussed about main important challenges that are facing in malware detection classification.

Index Terms: Malware, Malware Analysis, Static Analysis, Dynamic Analysis, Classification, Machine learning, Data mining Techniques, Malicious Code.

1 INTRODUCTION

Malware word defines from Malicious Software. Malware is a malicious code that affects the user system or computer and intently harms the computer by an attacker. Malware is variant forms which are a virus, Trojan, backdoor, rootkits, ransomware, worm, botnet, spyware, adware, keyloggers, etc., and there is a wide range of their families are existing and massively growing on the internet daily. According to the survey [1] conducted by AV-Test Institute, it registers that everyday 350,000 new malicious code and potentially unwanted applications. Each malicious one is classified with respect to their behavior and saved accordingly by this institute and gives the malware statistics in 2018 is 847.34m malicious code is found and recorded and registered. Some of the malware attacks in history are Melissa was a macro embedded with a word file. When the user opens it the macro will execute and resend the virus to the first 50 people in the user's address book. It was designed by David L. Smith in 1999. Likewise, several malware attacks in history namely, My Doom worm in 2004, Stuxnet in 2010, wannacry in 2017. In this paper, we presented the literature work of previously existing works of malware detection classification using machine learning algorithms. Section 2 covers about malware analysis and types of malware analysis. Section 3 is all about literature of malware detection classification using machine learning algorithms. Section 4 is discussion section and followed by Section 5 for conclusion.

2 MALWARE ANALYSIS

The analyzing behavior, functionality, and impact of malware samples on a user system defined as malware analysis. The analyzing of malware samples variants in different ways which are signature-based, behavior-based and memory-based malware analysis.

- *P HarshaLatha, Research Scholar, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India. E-mail: harsha17latha@gmail.com*
- *Dr. R Mohanasundaram, Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India. E-mail: mohanasundaramr@vit.ac.in*

2.1 Types of Malware Analysis

There are basically three types of Malware Analysis in general which are Static Malware Analysis, Dynamic Malware Analysis, and Memory Malware Analysis briefly in Fig. 1.

2.3 Static Malware Analysis

The process of detecting or examining the malicious code without executing it is defined as static malware analysis. It is a signature-based malware analysis. In static malware analysis, static features such as Metadata strings, code, and import libraries are extracted and used in the feature selection or feature extraction phase in the machine learning classification. Most probably the input file type of static malware analysis is should be of the type exe, DLL, documents, Assembly code, byte code, etc., from these file types static features are extracted as output. The tools used for static malware analysis are PEiD, ssdeep, pafish, Yara, strings, IDA Pro [2], OllyDbg [3], LordPE [4], OllyDump[5], etc.,

2.4 Dynamic Malware Analysis

The process of detecting the behavior and functionality of malicious code while executing it is defined as a Malware Analysis. Dynamic features are system calls, file activities, process activities, and network activities. The dynamic malware analysis tools are used to extract the dynamic features of malicious code, tools are CWSandbox [6], Anubis[7], comodo automated analysis, ThreatTrack etc. The monitoring tools for dynamic malware analysis are Process Explorer [12], Process Monitor [13], Capture-BAT [14] (for registry monitoring and file system monitoring), RegShot[15] (for detecting System change), Wireshark [16] (for network monitoring), Process Hacker replace [17] etc.

2.5 Memory Malware Analysis

The process of examining the malicious code after executing it is defined as memory malware analysis. The features for memory analysis are shared libraries, running processes, hooking detection, network connections, rootkit connection, hidden artifacts, code injection, etc. The tools for memory analysis are volatility, pin tools, Valgrind, etc.

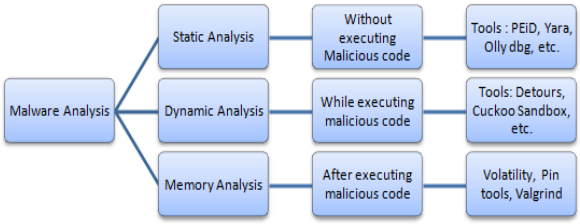

Fig. 1. *Types of Malware Analysis*

3 MALWARE DETECTION CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

Various Machine learning techniques are used for malware classification such as Support Vector Machine, Decision Tree, Naive Bayes, Random Forest, etc., and machine learning clustering techniques are used for clustering malware samples. The existing literature is discussed below about machine learning approaches for malware analysis. Gupta et al. 2018 [18] proposed architecture for malware detection. In this architecture, first prepares the dataset which contains 0.2 million of files in which has 0.05 million clean files and 0.15 million malware samples these are collected from different resources like VXheaven, Nothink, VirusShare, etc. After the collection of malware samples, automated malware analysis is performed on the collected dataset by using cuckoo sandbox which contains a series of python scripts to determine the behavior of malware during their execution. The results are in the format of JavaScript Object Notation (JSON). The static and dynamic features are extracted from the json reports using python in Apache Spark. After the feature extraction, the classification models are applied to it with 10-fold cross-validation. The classification models are applied to the dataset and evaluate the performance using the parameters True Positive Rate (TPR), False Positive Rate (FPR), Precision, False Negative Rate (FNR) and Accuracy. The experimental results show Naïve Bayes(NB), Support Vector Machine (SVM), and Random Forest (RF) gives 89.13 %, SVM gives 94.03% and Random Forest gives 98.88% of accuracy. Random Forest gives the best accuracy with minimum FPR and FNR. Cho et al. 2016 [19] propose a framework that performs preprocessing the data and classification (Malware Similarity). In the preprocessing step, the malware samples are reduced the amount in which it is categorized into malware families which is relevant to one another. In the classification process, behavior monitor, sequence refining, sequence alignment, and similarity calculation are to be addressed. 150 malware samples are used in 10 different malware variant families and the classification is repeated by 5 times and the experimental result gives 87% of high accuracy. Burnap et al. 2018 [20] present the novel approach in classification which uses self-organizing maps are used to distinguish between the malicious and benign files and it reduces the overfitting process when samples are training. Dataset is collected from VirusTotal API. In this architecture, multiple classifiers are used for classification and tested in different circumstances. Classifier models used in this approach are Random Forest, BayesNet, MLP, and Support Vector Machine. First, all the malware samples are tested under 10 fold cross-validation and the classifier results will give the best result by Random Forest with 98% accuracy and the same Random Forest accuracy is down by 12% when the Random Forest classifier is applied for different datasets due to overfitting problem. For this problem, they introduce the novel approach using Self Organizing Feature Map (SOFM) is a classifier that is applying the ANN approach. The experimental results are increased by 7% accuracy when compared to the previous model. Ab Razak et al. 2016 [21] presents the works of bibliometric analysis study of 4000 articles which are published from 2005 to 2015 from the ISI web of science database of records of the countries North America, Asia, and other continents research activities are discussed. With Search keywords of “malware”, they found 4546 records in a web of science database which includes journals, books, book sections, patents among that 2158 records of un-English articles are excluded which are KCI- Korean Journal Database, Derwent Innovation Index, and SCiELO Citation Index. The findings are collected from Impact journals, highly- cited articles, research areas, productivity, keywords frequency, institutions, and authors. They analyze and concluded that North America is crucial one of the publications of academic research in malware and Asia is in the second place of publishing malware articles. Ray et al. 2016 [22] give a brief overview of malware and malware analysis. They overviewed about different types of malware such as worm, Trojan horses, viruses, spyware, backdoor, and rootkits. And they described types of malware analysis which are static malware analysis and dynamic malware analysis. Without executing the malware comes under static analysis and with the execution of malware comes under dynamic analysis. Function call monitoring, function parameter analysis, information flow tracking, instruction tracing, etc., are the various techniques carried out in dynamic malware analysis. And some of the tools used for malware analysis are Norman sandbox, Anubis, CWSandbox, etc. They stated that dynamic malware analysis is a better method for malware analysis. Ucci et al. 2017 [23] present a survey on malware analysis through machine learning techniques. First, they present the review work of malware analysis objectives, features and machine learning algorithms to process malware analysis features. Second, they discussed the issues regarding Malware Datasets and they visualize the literature work of datasets from different sources. Third, they present the work regarding the novel approach of malware analysis economics, which mainly focused on performance metrics like accuracy, execution time and economic costs. AlAhmadi et al. 2018 [24] propose a novel approach to a malclassifier. It is a three-step process. In the first step, pre-processing and subsequent extraction is performed malware variant families are the input for this step which is from network traffic. These variant families are reassembled with network flow and then it goes to flow encoding, there subsequent extraction of the flow encoding is taken as the input. In the second step, the profile extraction performed on the input of sequence extraction, here the flow values are compared for similarity using binary similarity, Levenshtein distance, cosine similarity, interflow distance after that malware family n-flow mining is performed on the similarities, from their network behavior profiles are extracted which is the outcome of second step. In the third step, training and building a model on profile extraction features. Machine learning algorithms KNN and Random Forest are used in classification and training the model. Finally, the malclassifier achieves 95.5% (F-measure) of the high accuracy of malware family classification. Khan et al. 2017[25] introduce the framework for malware detecting

(unwanted signatures). Both remote analysis and local analysis of malware detection techniques are used in their framework. A file is checked whether it is malicious or benign with help of signatures. In the remote analysis, various antivirus software's is used to analyzing malicious executables and API. In the local analysis, Anti-virtual machine, anti-debugger, URL analysis, string analysis, and packing analysis are used. Ronen et al. 2018 [26] provide a standard benchmark dataset that was announced by Microsoft Malware Classification Challenge which is cited by several malware researchers, and it is serving in the Kaggle competition. The dataset consists of 9 different malware families with 0.5 terabytes of huge data having more than 20,000 malware samples in byte code cited by nearly 50 research papers. Ye et al. 2017 [27] presented a survey on malware detection using data mining techniques which is focused on intelligent malware detection approaches. They illustrate two stages which are feature extraction and classification/clustering as important stages in malware analysis and detection. They presented the research works from 2011 to 2016 and issues and challenges in the malware detection using data mining approaches. Wang et al. 2017 [28] introduce the implementation and design of sandbox, feature extractor, and the classifier. There are mainly three stages in their work which are collector, extractor, and classifier. Collector contains static analysis program and dynamic execution with the module PinFWSandbox which records the dynamic information and log file information and it passed to the extractor stage. Extractor performs both static feature extraction, dynamic instruction feature extraction and system called feature extraction. Finally, the classifier performs the action of combining all the classifier models such as single model classifier result, system call classifier result, dynamic immediate classifier result and dynamic opcode classifier result to give the better result of the f1 score which gives nearly close to 96%. Pai et al. 2017 [29] present malware classification using clustering algorithms. Static features are extracted from the opcode sequences and with their scores. These static features are used for malware classification using the clustering algorithms K-means, Expectation-Maximization, and Hidden Markov Models. Among the clustering algorithms, Expectation-Maximization gives better results of accuracy. This is the different approach for malware classification which includes machine learning clustering algorithms. Gupta et al. 2016 [30] present a framework using windows API call sequences. Five malware classes are classified which are Worm, Trojan-Downloader, Trojan-Spy, Trojan-Dropper and Backdoor among 2000 malware samples using API call sequence and fuzzy hashing based classification. Liu et al. [31] provide an approach to evaluate and classify unknown malware instances to cluster with respect to their families. To classify the malware instances shared nearest neighbor clustering algorithm. The experimental result gives 98.9% accuracy for known malware instances and 86.7% of unknown malware is correctly classified of new malware instances. Makandar et al. 2017 [32] give an overview of malware analysis and detection techniques with different types of malware families. Different methods or techniques are used to detect and analyze malware, one method is to visualize malware in the form of an image, grayscale image, etc. Different existing works related to visualization of malware families are given an overview of their work. Nari et al. 2013 [33] propose the automated malware classification based on network activity or behavior of malware. From the network traces (pcap file) behavior tree is generated and then features are extracted from it and perform the classification using different machine learning algorithms. Finally, it is found that the J48 classifier gives better accuracy results among different anti-virus program comparisons. Kosmidis et al. 2017 [34] provide an automated framework to classify the unknown malware samples using neural networks. Feature engineering used to extract features from the malimg dataset. Perceptron, decision tree, nearest centroid, stochastic gradient, multilayer perceptron, random forest algorithms are used to classify the unknown malware. Random Forest gives better average accuracy results and testing time also considered as a parameter. Gandotra et al. 2014 [35] provide the integrated framework to classify the unknown malware using the integrated feature set from both static and dynamic features to get a better classification of malware samples. The evaluation of feature extracted dataset and classification is done by using the four classifiers which are Multilayer perceptron, IB1, Decision Tree and random forest. In the experiment results, Random Forest gives high accuracy results with 99.58% of detecting unknown malware and classification. Islam et al. 2013 [36] introduced the framework earlier than Gandotra et al. [35], it is similar to an integrated framework of static and dynamic features and performs classification using classifiers to classify the unknown malware. Gandotra et al. 2014 [37] surveyed the malware analysis basically on classification and usage of machine learning in malware analysis. Makandar et al. 2015 [38] propose a classification of malware families using artificial neural networks. The malware binaries are converted to grayscale images and the images are resized. From the resized image sub-band filtering is applied to extract features and then feature vector is formed. To extract texture features Gabor wavelet and GIST descriptor are used. To classify the extracted features feed-forward back propagation neural networks are applied and the experiment result gives 96.35 % accuracy of unknown malware classification. Kruczkowski et al. 2014 [39] used the Support Vector Machine (SVM) to classify the malware samples using three validation techniques which are cross-validation, Leave – one- out and Random Sampling (RS). Among the three validation techniques, Random Sampling gives better results of accuracy 94. 98%.Nataraj et al. 2011 [40] presented a novel approach to classify the malware samples. Malware binaries are converted to an 8-bit vector and an 8-bit vector to a grayscale image. Image texture and feature vectors are used for classification. The Gabor wavelet and GIST descriptor are used in feature extraction. In the experimental results, 98% of accuracy is obtained. Tian et al. 2009 [41] present malware classification using printable strings which are in malware executables in an efficient way. Five classifiers are applied which are Naïve Bayes, Support vector machine, IB1, Random forest and Decision Tree on extracted features. The efficiency of all classifiers is improved using the AdaBoostM1 meta-classifier. In the experimental results of WEKA, Random forest and IB1 gives better results of the overall accuracy of 97%.Khammas et al. 2015 [42] give malware classification using the n-gram technique. For feature selection from dataset Principal Component Analysis (PCA) is used for better real-time results. Neural Network (NN), Decision Tree (J48), Support Vector Machine (SVM) and Naive Bayes (NB) classifiers are used for classification. In the experimental results, 97% accuracy gives by SVM. Devesa et al. 2010 [43] present automatic detection

of malware using behavior logs. Using the sandbox environment, Qemu for emulation and Wine for simulation for extracting features from behavior logs of malware samples. The four classifiers Naive Bayes, Random Forest, J48 and SMO applied to the extracted features to get better performance. Random forest classifier gives high accuracy of 96.2%.Lin et al. 2015 [44] proposed an approach for malware classification using effective features of data using feature selection and feature extraction methods. Feature selection is performed on the dataset using the MG TF –IDF method precisely selected features of subsets of the dataset. These précised features are examined under the feature extraction method PCA which gives accurate results by reducing the dimensions of the dataset. Support Vector Machine classifier is applied to the extracted features to get high accuracy and better performance. Dhammi et al. 2015 [45] propose an approach to classify malware samples and clustering them. The data is collected from the cuckoo sandbox environment and the data pre-processing is performed using different classifiers in the WEKA tool. Among the five classifiers of machine learning, LMT classifier gives a better accuracy performance of 98.3%. The classified malware samples are clustered using k-means clustering algorithm and give better results. Schultz et al. 2001 [46] were first introduced the method of detecting unknown malicious samples using data mining approaches. Static features are extracted from the PE Executables using LibBFD. The three data mining classifiers Naive Bayes, RIPPER, and Multi-Classifier System are applied to the extracted features. The Multi-Naive Bayes Strings gives a high average accuracy of 97.76%.The following Table 1 gives the existing literature in detailed as to what tools are used for their work, what machine learning algorithm used, what are internet sources for dataset collection, what are the parameter they considered to meet their goals and what are the future works they suggested are listed in the Table 1.

TABLE 1

SURVEYED PAPERS EXISTING LITERATURE WORK

Paper	*Tools Used*	*Algorithms used*	*Dataset Sources*	*Results*	*Future Work*
[18]	*Apache Spark, MLlib, Cuckoo Sandbox, Oracle VM VirtualBox*	*Naive Bayes, Random Forest, Support Vector Machine*	*VX Heaven, Nothink, VirusShare.*	*Among the three classifiers Naive Bayes, Support Vector Machine, and Random Forest, the Random Forest gives the best accuracy*	*Supervised Dimensionality Reduction in a large data set of malware samples.*
[19]	*Cuckoo Sandbox*	*Sequence alignment*	*VX Heaven*	*The average accuracy of all malware families is 87% and overall execution time is reduced from 91% to 99%*	*The pair wise sequence alignment algorithm will improve the performance proposed system for malware samples classification and execution time.*
[20]	*Cuckoo Sandbox*	*Random Forest, BN, MLP, Support Vector Machine, Self Organizing Feature Map*	*Virus total API.*	*Performance increased by 7.24% to 25.68% when compared to the previous model*	*Increasing the sample size and granularity of data for other models will perform better?*
[21]	*VOS viewer*	*Bibliometric analysis*	*Web of science*	*North America producing more research articles when compared to all other continents.*	*Finding the past 10 years of articles and analyze them with the keywords of "Malware Analysis", "Malware detection", "algorithm", "security" etc., which are important in research related to malware.*
[24]	-----	*KNN, Random Forest*	*CTU-13, Stratosphere IPS project*	*Performance is increased. 95.5% of high accuracy achieved.*	*Identifying the network behavioral changes in malware samples.*
[28]	*PinFWSandbox*	*The static immediate feature, system call feature, dynamic immediate feature, dynamic opcode*	*Vxheaven.org, malwr.com, viruscan.org*	*The merged model increases the percentage to 96%*	*It can apply the sequence feature or function sequence feature to get better results.*
[29]	-----	*Clustering algorithms*	*Collected from Malicia website*	*Expectation-Maximization techniques give a high accuracy of classification using*	*Apply the same technique with static features using call graphs and other measures and also*

				clustering algorithms.	apply for extracted dynamic features.
[30]	ssdeep	Windows API call sequence and fuzzy hashing.	VxVault, http://www.vxvault.net, Vxheaven, http://www.vxheaven.org VirusSign, http://www.virussign.com VirusTotal. https://www.virustotal.com	Five malware classes are successfully classified using windows API call sequence and fuzzy hashing based classification.	Text pattern matching techniques are used to classify the malware
[31]	---	Shared nearest neighbour (SNN).	Kingsoft, ESET NOD32, and Anubis.	Gives 98.9% accuracy of known malware and 86.7% of accuracy for unknown malware.	---
[33]	WEKA	J48, C4.5	Communication Research Center Canada (CRC).	Comparison between different antivirus programs gives better accuracy results	----
[34]	----	Decision tree, nearest centroid, perceptron, Stochastic gradient, Multilayer Perceptron, Random Forest	Malimg dataset.	Random Forest gives better accuracy results.	---
[35]	WEKA	MLP, DT, IB1, Random Forest	University of California	Random Forest gives high accuracy results of 99.58%	----
[36]	WEKA	Random Forest, IB1, DT, Support Vector Machine	CA's (Computer Associates) VETZoo	Random forest gives better accuracy and improves performance by 9% accuracy.	
[38]	Neural network	Feedforward back propagation neural network (ANN).	Mahenhur Dataset	The feedforward backpropagation neural network gives 96.35% of accuracy in experimental results.	Dimensionality reduction of the feature vector.
[39]	---	Support Vector Machine, cross-validation, leave-one-out, random sampling	N6 platform from Research and Academic Computer Network (NASK)	Support Vector Machine random sampling gives better accuracy results of 94.98%	Classification of a large dataset
[40]	Text editors and Binary editors	KNN	GIST image features	Using KNN 98% accuracy is obtained.	Clustering of malware samples using image-based features.
[41]	WEKA	Ada Boost, IB1, Random Forest, Support Vector Machine, DT, Naive Bayes	CA's VET zoo	Overall classification accuracy is 97%.	----
[42]	WEKA	Support Vector Machine, NN, J48, Naive Bayes, PCA	VX Heaven	The combination of feature selection techniques and Support Vector Machine classifier gives 97% of accuracy.	n-gram rule from SNORT signature for better accuracy using machine learning.
[43]	CWSandbox, Qemu (emulation) and Wine (simulation)	Naive Bayes, Random Forest, SMO, J48	VX Heaven	Random Forest gives high accuracy of 96.2%	Expanding features with including static analysis techniques.
[44]	Sandbo	Support	Institut	SVM	----

	x	*Vector Machine, PCA, MG TF-IDF*	*e of Information Industry*	*classifier with feature selection and feature extraction methods gives better performance and accuracy.*	
[45]	*WEKA*	*LMT, Support Vector Machine, Ridor, KNN, Naive Bayes, K-means*	*Online sources*	*LMT classifier gives a high accuracy of 98.28%. clustering of malware samples k-means gives better performance*	*Expand for large datasets and test for a combination of more classifiers to get better results.*
[46]	*---*	*Naive Bayes, RIPPER, MNB*	*FTP sites at Columbia University*	*Multi-Naive Bayes gives high accuracy of 97.76% of classifying malware*	*Utilization of bye sequence to extending work.*

4 DISCUSSION

Previous existing literature works of malware detection prove that successfully classification is done with the help of machine learning techniques but still there are some issues have not resolved. Zero-day attacks are the one which the day having no new malware will rise in future. It is the main aim of all the malware researchers. Basing on the survey [1] by AV-Test still, millions of new malware are rising day-by-day. Some issues and challenges of malware detection are still there and not yet resolved. One issue is to real verification or manual verification of classification results becomes harder when it comes to reality. Another issue is to improve the advancement of techniques for more active learning. The more recent advancements are expecting from the fields of machine learning, ensemble learning, deep learning and more. The most advanced techniques are needed to achieve Zero-day Attacks. Dealing with large datasets is also one of the issues. These advanced techniques are needed in dimensionality reduction.

5 CONCLUSION

This paper presents the survey about existing literature on malware analysis using different machine learning algorithms. Table 1 defines the different literature of existing works with what are the tools used in their work, what are the machine learning algorithms they used in their work, from what sources dataset is collected, what are parameters they consider to reach their goal and the corresponding experimental results and what are the future works are proposed all are listed in the table form. In the discussion, it clearly identifies that machine learning algorithms are very useful for the classification and clustering of malware samples for small datasets and for large volumes of data.

6 REFERENCES

[1] AV-TEST (2018, November 28). The Independent IT-Security Institute, Malware Statistics [Online]. Available: https://www.av-test.org/en/statistics/malware/

[2] IDAPro. (2018, November 28). [Online]. Available: https://www.hex-rays.com/products/ida/support/download_freeware.shtml

[3] OllyDbg. (2018, November 28). [Online]. Available: http://www.ollydbg.de/

[4] LordPE. (2018, November 28). [Online]. Available: http://www.woodmann.com/collaborative/tools/index.php/LordPE

[5] OllyDump. (2018, November 28). [Online]. Available: http://www.woodmann.com/collaborative/tools/index.php/OllyDump

[6] Willems, C., Holz, T. and Freiling, F. (2007) Toward Automated Dynamic Malware Analysis Using Cwsandbox. IEEE Security & Privacy, 5, 32-39. http://dx.doi.org/10.1109/MSP.2007.45

[7] Anubis. (2018, November 28). [Online]. Available: http://anubis.iseclab.org/

[8] Bayer, U., Kruegel, C. and Kirda, E. (2006) TTAnalyze: A Tool for Analyzing Malware. Proceedings of the 15th European Institute for Computer Antivirus Research Annual Conference.

[9] Norman Sandbox. (2018, November 28). [Online]. Available: http://sandbox.norman.no

[10] Dinaburg, A., Royal, P., Sharif, M. and Lee, W. (2008) Ether: Malware Analysis via Hardware Virtualization Extensions. Proceedings of the 15th ACM Conference on Computer and Communications Security, CCS'08, Alexandria, 27-31 October 2008, 51-62.

[11] ThreatExpert. (2018, November 28). [Online]. Available: http://www.threatexpert.com/submit.aspx

[12] Process Explorer. (2014). [Online]. Available: http://technet.microsoft.com/en-us/sysinternals/bb896653.aspx

[13] Process Monitor. (2014). [Online]. Available: http://technet.microsoft.com/en-us/sysinternals/bb896645.aspx

[14] Capture BAT. (2018, November 28). [Online]. Available: https://www.honeynet.org/node/315

[15] Regshot. (2018, November 28). [Online]. Available: http://sourceforge.net/projects/regshot/

[16] Wireshark. (2018, November 28). [Online]. Available: http://www.wireshark.org/

[17] Process Hacker replace. (2018, November 28). [Online]. Available: http://processhacker.sourceforge.net/

[18] Gupta, D., & Rani, R. (2018). Big Data Framework for Zero-Day Malware Detection. Cybernetics and Systems, 49(2), 103-121.

[19] Cho, I. K., Kim, T. G., Shim, Y. J., Ryu, M., & Im, E. G. (2016). Malware Analysis and Classification Using Sequence Alignments. Intelligent Automation & Soft Computing, 22(3), 371-377.

[20] Burnap, P., French, R., Turner, F., & Jones, K. (2018). Malware classification using self organising feature maps and machine activity data. computers & security, 73, 399-

410.
[21] Ab Razak, M. F., Anuar, N. B., Salleh, R., & Firdaus, A. (2016). The rise of "malware": Bibliometric analysis of malware study. Journal of Network and Computer Applications, 75, 58-76.
[22] Ray, A., & Nath, A. (2016). Introduction to Malware and Malware Analysis: A brief overview. International Journal, 4(10).
[23] Ucci, D., Aniello, L., & Baldoni, R. (2017). Survey on the usage of machine learning techniques for malware analysis. arXiv preprint arXiv:1710.08189.
[24] AlAhmadi, B. A., & Martinovic, I. (2018, May). MalClassifier: Malware family classification using network flow sequence behaviour. In APWG Symposium on Electronic Crime Research (eCrime), 2018 (pp. 1-13). IEEE.
[25] Khan, M. H., & Khan, I. R. (2017). Malware Detection and Analysis. International Journal of Advanced Research in Computer Science, 8(5).
[26] Ronen, R., Radu, M., Feuerstein, C., Yom-Tov, E., & Ahmadi, M. (2018). Microsoft Malware Classification Challenge. arXiv preprint arXiv:1802.10135.
[27] Ye, Y., Li, T., Adjeroh, D., & Iyengar, S. S. (2017). A survey on malware detection using data mining techniques. ACM Computing Surveys (CSUR), 50(3), 41.
[28] Wang, C., Ding, J., Guo, T., & Cui, B. (2017, November). A Malware Detection Method Based on Sandbox, Binary Instrumentation and Multidimensional Feature Extraction. In International Conference on Broadband and Wireless Computing, Communication and Applications (pp. 427-438). Springer, Cham.
[29] Pai, S., Di Troia, F., Visaggio, C. A., Austin, T. H., & Stamp, M. (2017). Clustering for malware classification. Journal of Computer Virology and Hacking Techniques, 13(2), 95-107.
[30] Gupta, S., Sharma, H., & Kaur, S. (2016, December). Malware Characterization Using Windows API Call Sequences. In International Conference on Security, Privacy, and Applied Cryptography Engineering (pp. 271-280). Springer, Cham.
[31] Liu, L., Wang, B. S., Yu, B., & Zhong, Q. X. (2017). Automatic malware classification and new malware detection using machine learning. Frontiers of Information Technology & Electronic Engineering, 18(9), 1336-1347.
[32] Makandar, A., & Patrot, A. (2015). Overview of malware analysis and detection. In IJCA proceedings on national conference on knowledge, innovation in technology and engineering, NCKITE (Vol. 1, pp. 35-40).
[33] Nari, S., & Ghorbani, A. A. (2013, January). Automated malware classification based on network behavior. In 2013 International Conference on Computing, Networking and Communications (ICNC) (pp. 642-647). IEEE.
[34] Kosmidis, K., & Kalloniatis, C. (2017, September). Machine Learning and Images for Malware Detection and Classification. In Proceedings of the 21st Pan-Hellenic Conference on Informatics (p. 5). ACM.
[35] Gandotra, E., Bansal, D., & Sofat, S. (2014, September). Integrated framework for classification of malwares. In Proceedings of the 7th International Conference on Security of Information and Networks (p. 417). ACM.
[36] Islam, R., Tian, R., Batten, L. M., & Versteeg, S. (2013). Classification of malware based on integrated static and dynamic features. Journal of Network and Computer Applications, 36(2), 646-656.
[37] Gandotra, E., Bansal, D., & Sofat, S. (2014). Malware analysis and classification: A survey. Journal of Information Security, 5(02), 56.
[38] Makandar, A., & Patrot, A. (2015, December). Malware analysis and classification using Artificial Neural Network. In Trends in Automation, Communications and Computing Technology (I-TACT-15), 2015 International Conference on (pp. 1-6). IEEE.
[39] Kruczkowski, M., & Szynkiewicz, E. N. (2014, August). Support vector machine for malware analysis and classification. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02 (pp. 415-420). IEEE Computer Society.
[40] Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. S. (2011, July). Malware images: visualization and automatic classification. In Proceedings of the 8th international symposium on visualization for cyber security (p. 4). ACM.
[41] Tian, R., Batten, L., Islam, R., & Versteeg, S. (2009, October). An automated classification system based on the strings of trojan and virus families. In Malicious and Unwanted Software (MALWARE), 2009 4th International Conference on (pp. 23-30). IEEE.
[42] Khammas, B. M., Monemi, A., Bassi, J. S., Ismail, I., Nor, S. M., & Marsono, M. N. (2015). Feature selection and machine learning classification for malware detection. Jurnal Teknologi, 77(1).
[43] Devesa, J., Santos, I., Cantero, X., Penya, Y. K., & Bringas, P. G. (2010). Automatic Behaviour-based Analysis and Classification System for Malware Detection. ICEIS (2), 2, 395-399.
[44] Lin, C. T., Wang, N. J., Xiao, H., & Eckert, C. (2015). Feature Selection and Extraction for Malware Classification. J. Inf. Sci. Eng., 31(3), 965-992.
[45] Dhammi, A., & Singh, M. (2015, August). Behavior analysis of malware using machine learning. In Contemporary Computing (IC3), 2015 Eighth International Conference on (pp. 481-486). IEEE.
[46] Schultz, M. G., Eskin, E., Zadok, F., & Stolfo, S. J. (2001). Data mining methods for detection of new malicious executables. In Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on (pp. 38-49). IEEE.