ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Different Machine Learning Techniques for Intrusion Detection System in Network Environment

Divyarani P. Babar, Dr. Pankaj Agarkar

P.G. Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engg, Pune, India

Assistant Professor, Department of Computer Engineering, Dr. D.Y Patil School of Engg, Pune, India

ABSTRACT: Intrusion Detection System (IDS) has emerged as one of the most complicated malware which encrypts the files once it enters into IDS the system. For the last couple of years, the rise in the number of IDS attacks with increasing severity, potency to cause high damage, ease of carrying out the attack has caused the conventional anti-malware techniques to pay attention and include advanced IDS detection mechanism. The IDS families, collectively characterized as crypto IDS, encode certain file types on infected systems and forces users to pay the ransom through certain online payment methods to get the key for decryption. Hindering of IDS is essential in order to hoard the user's file. This paper proposes an approach by analyzing the program binaries both statically and dynamically to find specific properties for IDS. Our approach has identified such static IDS-specific properties along with the run-time properties, typically present during IDS execution that can be used for classification of IDS. The proposed system illustrates the approach of supervised learning which generate some Background Knowledge (BK) during the training phase and apply the same in testing phase.

KEYWORDS: Machine Learning, IDS attack, Network attacks, supervised learning

I. INTRODUCTION

IDS has been widely used to target the computers recently, and new types are being constantly added to the family. IDS reaches the target computer using social engineering techniques and activates after the victim downloads the email attachments or clicks on insecure links. As it spreads over the network easily, the number of victims exposed to this threat is very high.

Since IDS hides its location once it reaches the target through jump points, detecting an attack becomes extremely difficult. The IDS encrypts the most commonly used files in the target computer; therefore, individual users or even institutions are forced to pay a ransom for decryption. Attackers nowadays prefer Bitcoin, an anonymous payment mechanism, as a means for payment, thereby concealing their identity and location. These peculiarities encourage attackers to choose IDS, compared to other malware, in their malicious attacks.

Malware are of various kinds based on their behaviour, properties and propagation technique. IDS is a type of malware with specific properties and objectives. It can lock the system or can encrypt the data as well. An attacker can demand a desired amount of money in return of decryption key. The advanced cryptographic algorithms are devised to protect valuable information and data, yet on the other side, these algorithms are being used to create malicious programs, specifically 'IDS'

The IDS use encryption algorithms that vary from RC4 to RSA+AES and ECDH+AES. While the attackers used only symmetric encryption methods with the advent of the IDS threat, they now prefer the hybrid encryption mechanism, which utilizes both symmetric and asymmetric encryption methods [18].

In the 6 new generation IDS attacks, only the public key is sent to the infected machine, from the cryptographic key pair created on the command and control server, and the confidential key is never let out from the server. Thus, after encoding on machines exposed to today's IDS attacks is completely ended, the private key stored in the C&C server is needed to open the files [18].

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

This paper focuses on studying the IDS structure and behavior to retrieve its specific properties. The objective is to identify these as features for IDS classification using machine learning techniques and anomaly-based detection.

IDS, with is continuously improving attack policies and its spread in the network as well as computing environment has now become most common type of malware used by many attackers to get unauthorized access of victim machines. Most of the antivirus software using signature-based detection methods fail to detect the malicious activities because they perform analysis using hash signature or rules which is already stored in databases. Because hash signature samples of zero-day attacks do not exist in antivirus software databases, identify the IDS by using anomaly-based detection method is more efficient than the signature-based system. The most important aspect for anomaly-based recognition method is to be able to analyze the data adequately to determine the ability and the behavior of the IDS. This approach consists of two segments. These are "training" and "testing" sections. In the training phase, we have used rule creation techniques using some learning function called as supervised learning approach, using this learning function and some background knowledge has considered for creating the learning rules of the system. That is, machine learning algorithms have not been used. In second phase system, we evaluate the network stream data for any file metadata has evaluated with learning rules using some pattern matching algorithms, and categorized the current file is normal or malicious according to similarity weight with signatures or rules.

II. LITERATURE SURVEY

Many solutions have been proposed till now regarding of malware detection classification but new kind of malware are so clever that they can hide them self and some of them are very difficult to detect like evasive malware, IDS, and backdoors. We have carried a thorough study of the advanced malware detection approaches proposed by multiple researchers. The brief summary of a few major approaches is given below:
Mike Wojnowicz et al: has explained in [1]: Structural Entropy Analysis for Automated Malware Classification: they describe a technique called structural entropy analysis which can help to detect the presence of suspicious obfuscation and meta-obfuscation techniques. Structural entropy analysis goes beyond mean entropy analysis and mere "detection of packing" to find suspicious patterns of entropy change.

Igor Kononenko et al:[2] : has explained about Semi –naïve Bayesian classifier: in this paper, the algorithm of the 'naïve Bayesian classifier that assumes the independence of attributes is extended to detect the dependencies between attributes. The idea is to optimize the trade-off between the 'non-naivety' and the reliability of approximation s of probabilities. Experiments in four medical diagnostic problems are described. In two domains whereby the expert opinion the attributes are infected independent, the semi-naïve Bayesian classifier achieved the same accuracy as naïve Bayes. In two other domains, the semi-naïve Bayesian classifier slightly outperformed the naïve Bayesian classifier.
Charles LeDoux et al: [3]: in 2015 told about the Malware contains inherent patterns and similarities due to code and code pattern reuse by malware authors. Machine learning operates by discovering inherent patterns and similarities. In this paper, the concept of machine learning and how it is being applied in malware analysis has been described.

Loan Pop at al: [6] has explained the implementation of the Naïve Bayes (shortly NB) classifier. It shows that the algorithm NB improves the tasks of the Web Mining by the accuracy documents Classification. Its applications are important in the following areas: E-mail spamming; filtering spam results out of search queries; mining Log files for computing system management; machine learning for Semantic Web; document ranking by text classification; hierarchical text categorization; managing content with automatic classification and other areas from Web Mining.

Mehdi, B. et al: [7] their approach is using Variable length sequence of system call rather than fixed length. This concept generalized the n-gram approach known as hyper grams Where k-dimensional hyperspace is envisioned having variable length N-gram in which process moves depending on its own sequence. The process marks its impact in this space and generates hyper-grams that act as its profile while classification. The results are given by the author showing better performance than traditional n-gram.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Shahzad, F., Bhatti [8], [9] has given an approach to detect the run-time behavior of a process known as genetic footprint was proposed. Information maintained in the process control block (PCB) of an executing process was used as features. The author also performed a comparison with other existing solutions.

EktaGandotra et al: [11] gave a survey about malware analysis and classification. They pointed out the challenges in malware detection because of polymorphism and metamorphism techniques used by malware. The behavioral patterns can be extracted statically or dynamically can be used to classify the malware.

There are certain approaches suggested which focus solely on IDS detection:
Amin Kharraz et al:[10] explained about IDS attacks that have been observed in the wild between 2006 and 2014. They analyzed 1359 samples that belong to 15 different IDS families. They explained the IDS encryption, deletion and communication techniques. Examination of file system activities of IDS samples suggested that it is possible to detect IDS by looking at I/O request and Master File table in NFTS file system.

SatnamSingh et al:[12] has given an approach for detection of IDS using command and control server DNS logs. IDS uses DGA algorithm to generate random fake domain name by detecting the encryption key while communication with command and control server, so that detection of a valid or invalid domain name is difficult. Cryptolocker, WanaCry, TeslaCrypt use DGA for generating domain name.

RehmanArshad and Kung-Kiu-Lau et al: [17] have explained the extraction of architecture from Legacy code using Static reverse engineering techniques.

The existing work on the IDS is focused upon analyzing the working mechanism of IDS. However, the suitability of conventional malware analysis and detection mechanisms for IDS is yet to be assessed. This paper aims to recognize the exclusive properties of IDS using static and runtime malware analysis techniques.

III. IDS DETECTION METHODOLOGY

This section discusses the application of existing malware detection techniques to IDS and then the proposed approach. There are two different methods are available to detect malware in the system, "signature-based detection", and "anomaly-based detection" methods [20] like supervised base machine learning systems. Each of the methods can be applied using one of the three techniques - Static approach, dynamic approach and hybrid approach [20]. Figure 2 shows these detection methods and techniques with respect to IDS.

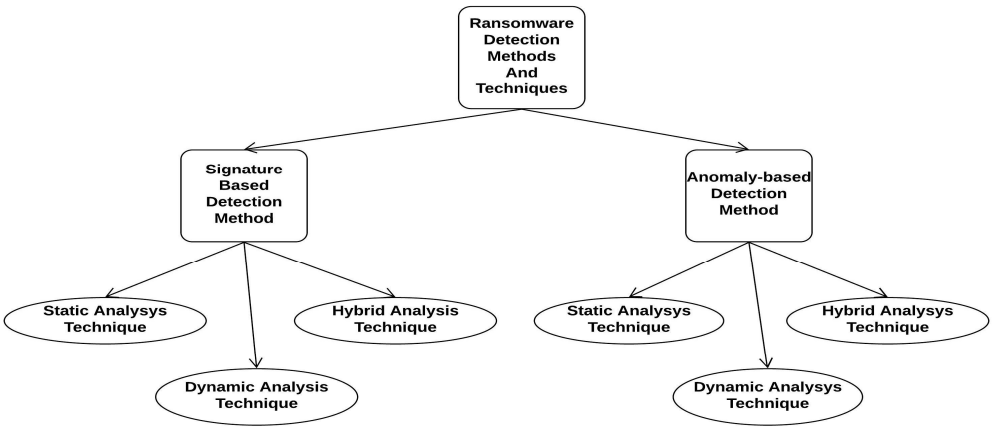

Figure 2: IDS Detection Methods and Techniques [20]

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

3.1 IDS Detection Techniques

Three different strategies have been used in many existing approaches, a static method is the first version of supervised learning and dynamic is a bit advanced than. The hybrid method is basically a combination of static as well as dynamic method which works like reinforcement learning approach, each of detection methods is explained below.

Static analysis: In this technique, the syntax or structural features of malware are used. This technique is used for detection before malware is run [20]. In IDS systems when packets arrive on victim's port, the system captures it and store in own buffer class. The buffer class does not allow entry to such packets in machine or device. During this phase, the system generates hash values for the received packet and maps the features similarity with database signatures and defines the current file or packet is malicious or normal.

Dynamic analysis: In this technique, runtime information of malware is used. This technique is used for detection while malware is running or afterward [20]. During the execution it generates the runtime hash of current metadata and evaluates with the database pattern, the similarity weight of current hash verified with quality threshold values and assign the class label to the current test data connection.

Hybrid analysis: This technique consists of the use of the features of two techniques which is already mentioned. This technique is used for detection before, during, and after the malware is run [20]. It has superior benefits over other approach. The proposed system works with similar detection approach which is already used in some existing machine learning base IDS.

3.2 IDS Detection Methods

Learning system basically work with fold validation and cross validation approach, in each technique system creates some signature based as well as anomaly base detection rules or policies. The same strategy ss carried out in two different methods to detect IDS, as explained below.

Signature-based detection: In this method, IDS is detected by signature information. Storage area such as repository is needed for storing signatures [20]. Current signature-based IDS detection systems are largely based on syntactic signatures. These systems are equipped with a database of regular expressions that are considered malicious [21]. Static and hybrid signature-based detection methods are shown in figure 3 and figure 4 both strategies works on runtime dynamic environments.

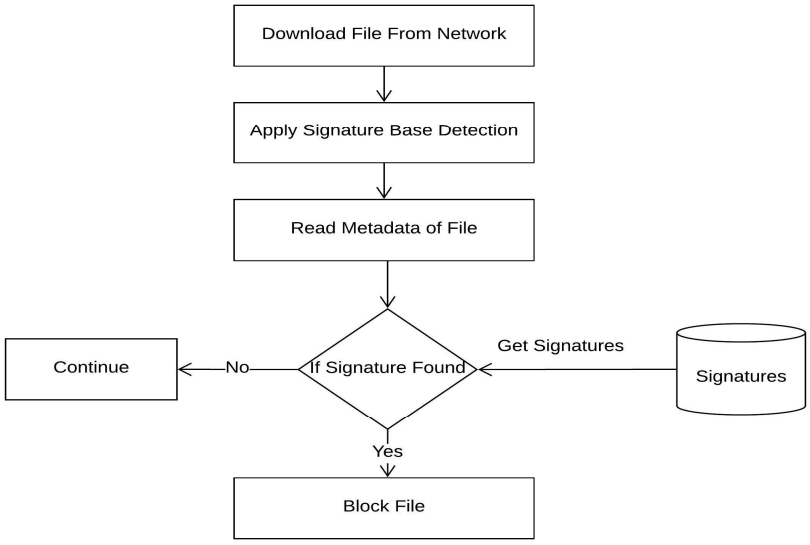

Figure 3: Static Signature base detection

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

The figure 3 and 4 illustrates when the signature of the IDS is created, it is recorded in the signature repository. The main reasons are to reduce the number of IDS signatures in the repository, and speed up the subsequent detection rate [20].

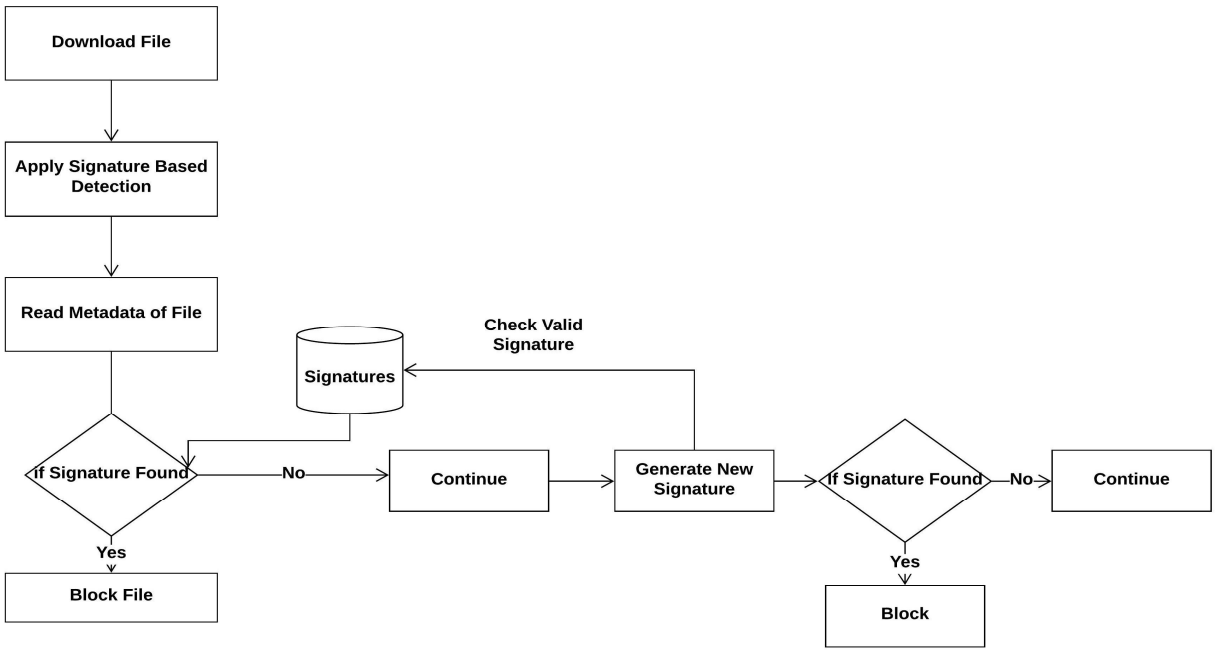

Figure 4: dynamic Signature base detection

Static signature-based detection systems only detect IDS whose signatures are already stored in the database. These systems fail to detect new, previously unseen threats, that is, zero-day attacks. This method is also less susceptible to obfuscation techniques [20, 21]. These systems can be easily deceived since signatures stored in databases are only based on regular expressions and string matching.

Anomaly-based detection: In this method, it is essential to have the knowledge to decide whether the software is harmful. Rule sets are used to determine whether the software has established benign and valid behavior [20]. The major drawback of this method is defining its rule sets [22]. But the anomaly-based detection systems work steadily once the rule sets are well defined.
The main advantage of this method compared to signature-based detection systems is that it is possible to detect zero-day attacks that are not known before [22]. So, benign or malicious behavior can be detected, whether it is a known attack or not [23]. This method consists of two phases. These are “training/learning” and “monitoring/detecting” phases. During the training phase, the benign behaviors of the system are revealed and the patterns of the system are observed [20, 23] After the benign behaviors are defined in the training phase, this method switches to “monitoring” phase and compares the real-time system with benign behaviors learned during the previous phase [23].

The monitoring phase generates an alert or warning message when something excepting benign malicious behaviors occurs in the system [20]. Disadvantages of this method are that it has high false-positive, too many alerts, and also difficulty and complexity in determining rule sets during the training phase [20]. Anomaly-based detection methods are shown in Figure 5.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

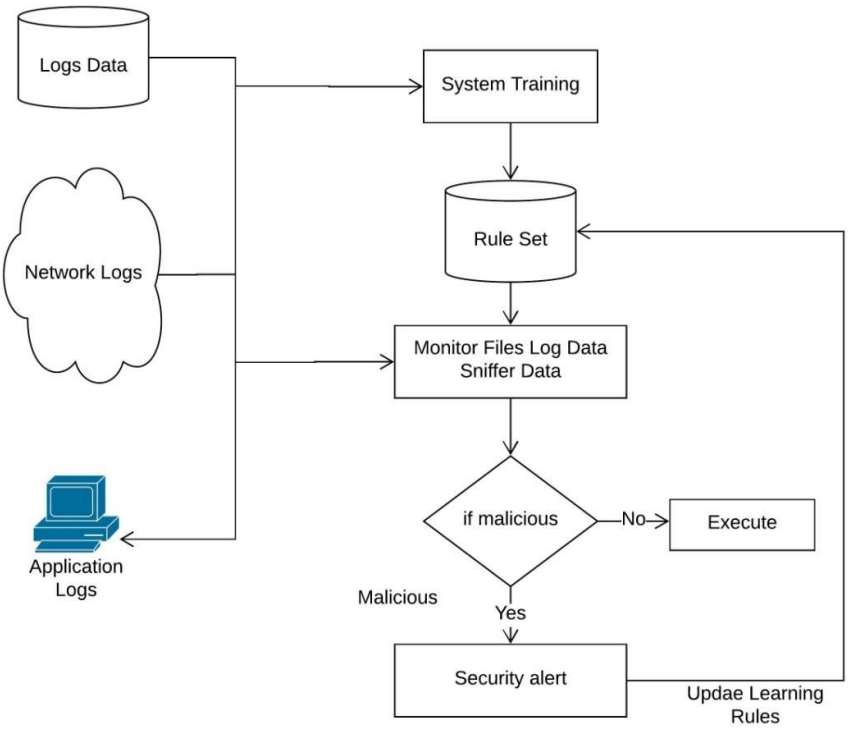

Figure 5 :Anomalybased detection method

3.3 Proposed System Architecture

The proposed system illustrates how to identify the various attacks features generated from an attacker on target victim machine and the system develops itself constantly against the security measures taken. The IDS continuously adds new features beside the security methods. Today, considering the situation of IDS they reached, security solutions like firewalls and antivirus, which are classical detection methods using threat signature rules or policies in databases, fail to detect and eliminate the IDS. Therefore, using both static and behavioral methods to detect and prevent IDS is a wiser solution. Figure 6 shows the system architecture for runtime anomaly based IDS detection. The first section is training phase, which learns the rules according to background knowledge of system, stores them into the rule repository. This repository is used during the detection phase. An optimization algorithm is used for creating the dynamic rules in training module and pattern matching techniques like Cosine Similarity (CS) are used in testing phase.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

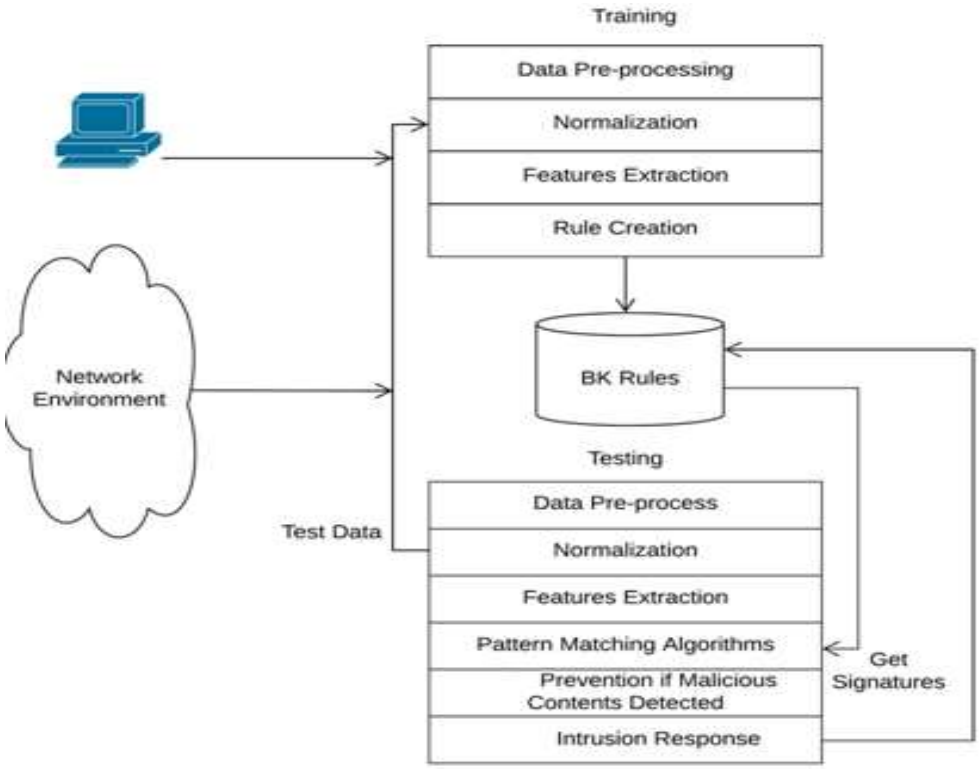

Figure 6 : Proposed system overview

3.3.1 Features for IDS

The rules are based on the features extracted from the PE file's header analysis. It is an offline analysis of an executable file, in order to study the structure of the file and extract static properties present in multiple sections of the file. The DOS header, file section header, and optional headers are analyzed to identify the prominent features for IDS, listed in Table 1. These IDS specific features are used for rules creation during the training phase. an

III. RESULTS AND DISCUSSIONS

Malware and IDS sample set is taken from VirusShare repository in PE format. The system has also used some Intrusion Detection Systems (IDS) dataset like KDDCup99, NSLKDD 2003, ISCX 2012, Botnet 2012 and WSNTrace simulation log files.

These datasets are used to generate the training rules for IDS in a vulnerable network environment using various machine learning algorithms. The prominent features selected in the IDS are as given in Tables 1 Each attributes gives the information of features of current packet received from a network stream.

Table 1: PE Features for IDS

Feature	Description
e_magic	Magic number
e_cblp	Bytes on the last page
e_sp	Initial SP value
e_csum	Checksum
e_ip	Initial IP value

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

e_lfanew	File address of the new exe header
Signature	The PE signature
Number of sections	Indicates the size of the section table
TimeDateStamp	Time and Date
Size of the code	The size of the code section, in bytes, or the sum of all such sections if there are multiple code sections.
Size of the initialized data	The size of the initialized data section, in bytes, or the sum of all such sections if there are multiple initialized data sections.
Ispacker	If packer is present then true otherwise false
Packer type	Type of packer used
Open mutex / Create mutex	Used for isolation
Entropy (e_file, e_text, e_data)	Entropy for file, text and data
Strings	Presence of some specific strings (Crypt/Decrypt/%amount%)
Ws2_32.dll	dll used for network connections and communication
Get_news	Command used to modify the registry
Add_entry	Store information from the client
Get_add	Get access to the file from the given path

The Table 3 illustrate network packet features for detect the NIDS and HIDS attacks.

Table 3: Packet features attributes and information

having_IP_Address	ip address available or not packet header
Pro	Protocol used
URL_Length	count the characters of URL length
shortining_service	URL may be made substantially shorter and still direct to the required page.
having_at_symbol	packet contains @ symbol
double_slash_redirecting	packet contains // or \\
Prefix_Suffix	URL contains some hide like /URL or /URL....
having_sub_domain SSLFinal_state	URL contain some domain SSL apply or not
Domain_Registration_Length	length of registered domain in URL
Favicon	Show icon for website in browser
HTTPS_token	HTTPS has used or not
Request_URL	requested URL
URL_of_Anchor	Anchor tag in URL
Links_in_Tags	Links available in tags

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

SFH	fishing page detetcted or not
Subimitting_to_email	submitting to email or not
abnormal_URL	is it abnormal url
Redirect	redirected url yes or no
on_mouseover	event generated on mouse over
rightclick	event generated on right click
popupwindow	event generated on pop up window
Iframe	no of frmaes contains in HTML page
age_of_domain	age of domain in years
DNSrecord	Record is stored in DNS server
web_traffic	web traffic is available or not
page_rank	page is ranked or not
google_index	page google index is true or not
link_to_pointing_page	is there any links which poot another pages

IV. CONCLUSION

The proposed system can be considered an IDS, illustrating intrusion prevention as well as intrusion response in a vulnerable environment. Using prevention approach, the system can recover the affected file or unblock the access control. The web link which contains the IDS features has present, it can be recognized and avoided. This helps in avoiding the early revealing of IDS and it can be prohibited by entering into the system. By using the proposed learning rules to extract the metadata from given files and compare with BK rules, it is easy to identify where IDS is created. The responsive approach can be used to eliminate future attacks with same type of signatures. Multiple experimental analyses illustrate how the proposed system provides better accuracy in detection phase where system recognized malicious activities. To build the different learning rules using various network security dataset and improve the detection accuracy is interested work for enhancement of system.

REFERENCES

[1] Mike Wojnowicz, Glenn Chisholm & Matt Wolf, “Structural Entropy analysis for Automated Malware Classification”, RSA conference, 2015.
[2] Kononenko, Igor. "Semi-naive Bayesian classifier." European Working Session on learning. Springer, Berlin, Heidelberg, 1991.
[3] LeDoux, Charles, and ArunLakhotia. "Malware and machine learning."Intelligent methods for Cyber Warfare.Springer, Cham, 2015.1-42.
[4] Lin, Chih-Ta, et al. "Feature Selection and Extraction for Malware Classification." J. Inf. sci.Eng. 31.3, 2015.965-992.
[5] Paul, Calvin B. Entropy-based file type identification and partitioning. Diss-monterey, California: Naval Postgraduate School, 2017.
[6] Pop, Ioan. "An approach of the Naive Bayes classifier for the document classification." General Mathematics 14.4 (2006): 135-138.
[7] Mehdi, Bilal, et al. "Towards a theory of generalizing system call representation for inexecution malware detection." Communications (ICC), 2010 IEEE International Conference on.IEEE, 2010.
[8] Shahzad, Farrukh, et al. "In-execution malware detection using task structures of linux processes." Communications (ICC), 2011 IEEE International Conference on.IEEE, 2011.
[9] Shahzad, Farrukh, Muhammad Shahzad, and MuddassarFarooq. "In-execution dynamic malware analysis and detection by mining information in process control blocks of Linux OS." Information Sciences 231 (2013): 45-63.
[10] Kramer, Simon, and Julian C. Bradfield. "A general definition of malware." Journal in computer virology 6.2 (2010): 105-114..
[11] Gandotra, Ekta, DivyaBansal, and SanjeevSofat. "Malware analysis and classification: A survey." Journal of Information Security 5.02 (2014): 56.
[12] “IDS Command and Control Detection Using Machine Learning.” Acalvio, 29 Mar. 2018, blog.acalvio.com/IDS-command-and-control-detection-using-Machine-learning.
[13] Crowe, Jonathan. “Must-Know IDS Statistics 2017.” Barkly Endpoint Security Blog, blog.barkly.com/IDS-statistics-2017.
[14] Ablon, Lillian, Martin C. Libicki, and Andrea A. Golay. Markets for cybercrime tools and stolen data: Hackers' bazaar. Rand Corporation, 2014.
[15] Egele, Manuel, et al. "A survey on automated dynamic malware-analysis techniques and Tools " ACM computing surveys (CSUR) 44.2 (2012): 6.
[16] Ravula, Ravindar Reddy. Classification of Malware: Using Reverse Engineering and machine Learning Techniques. LAP Lambert Academic Publishing, 2011.
[17] Burd, Elizabeth, Malcolm Munro, and ClazienWezeman. "Extracting reusable modules from legacy code: Considering the issues of module granularity." Reverse Engineering, 1996, Proceedings of the Third Working Conference on.IEEE, 1996.
[18] C. Krzysztof and M. Wojciech (2016). Using Software-Defined Networking for IDS Mitigation: The Case of CryptoWall, Network Forensics And Surveillance For Emerging Networks.
[19] Calyptix (2015). Critroni IDS Decryption: Not an Option, <https://www.calyptix.com/malware/critroni-IDS-decryptionnot-an-option/>, data retrieved 02.01.2018.
[20] N. Idika, A.P.Mathur (2007). A Survey of Malware Detection Techniques, Department of Computer Science, Purdue University.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

[21] Moser, C. Kruegel, and E.Kirda (2007). Limits of Static Analysis for Malware Detection, 23rd Annual Computer Security Applications Conference, 1063-9527/07 $25.00 © 2007 IEEE, DOI 10.1109/ACSAC.2007.21
[22] V.Jyothsna, V.V.R. Prasad, K.M.Prasad (2011). A Review of Anomaly based Intrusison Detection Systems, International Journal of Computer Applications (0975–8887), Volume 28-No.7, August 2011.
[23] W.Robertson, G.Vigna, C.Kruegel, and R.A.Kemmerer (2006). Using Generalization and Characterization Techniques in the Anomalybased Detection ofWeb Attacks, Proceedings of the Network and Distributed System Security Symposium, NDSS 2006, San Diego, California, USA
[24]MariemBelhor, Farah Jemili, "Intrusion Detection based on genetic fuzzy classification system", 2016 IEEE 13th International Conference on Computer Systems and Application (AICCSA), Nov 29 2016-Dec 2, 2016, Sousse, Tunisia.
[25] Greensmith J, Aickelin U. Firewalls, Intrusion Detection and Anti-virus Scanners. Computer Science Technical Report No. NOTTCS-TR-2005-1].Retrieved from http://citeseerx.ist. psu. edu/viewdoc/download. 2005.