A Survey on Methods to Detect Unknown Web Attacks

Divya L[1], Rajendra M[2]
[1]Student, Atria institute of technology, Bangalore, India
[2]Assistant professor, Atria institute of technology, Bangalore, India

Abstract- *Due to increasing volume and importance of web communication throughout the internet, there is a greater need for security protection. When protecting systems, security experts maintain a database containing featuring signatures of a large number of attacks for attack detection. This can detect only known attacks. There are two main types of detection techniques: signature-based ad anomaly based. Signature based detects only known attacks as unknown attack signature has not been written. Whereas, anomaly based can detect unknown attacks but the problem is that if the malicious traffic falls under normal pattern it is not detected. In this paper, we introduce various ways to detect the unknown attacks automatically by applying anomaly based intrusion detection system with other algorithms.*

Keywords- Intrusion detection system, anomaly detection system, classification, unknown web attacks

I. INTRODUCTION

Both home and enterprise networks need an effective security defence because of increasing number and frequency use of web-based applications. A comprehensive analysis carried out by Symantec [1] reveals that nearly one million threats are released into public network each day. A recent massive cyber-attack took place on 12th May 2017 and major impact in a significant element of the UK's National Health Service (NHS), other health industries and created chaos in hospitals across England. Thousands of computers at hospitals and GPs surgeries became victims of global ransomware attacks, derivatives of the WannaCry attack, which are believed to have exploited a vulnerability first discovered by the National Security Agency (NSA) [2].

In particular, the attack exploited a vulnerability in the Windows Server Message Block (SMB) protocol and installed backdoor tools to deliver and run a WannaCry ransomware package.

The more recent attack on September 2018 [3] was specifically designed to target Facebook the social network company. The hackers gained access to personal information of nearly 50 million users. The attackers gained access to user accounts and potentially took control of them by exploiting a feature in Facebook's code. The hackers used the flaws in Facebook system to break into users accounts, including Mark Zuckerberg, chief executive officer of Facebook and Sheryl Sandberg, chief operating officer of Facebook. The first two flaws were introduced by an online tool meant to improve the privacy of users and third flaw was introduced in July 2017 by a tool meant to easily upload birthday videos. The hackers instead of going for payment information or passwords, they stole personal information such as names, relationship status, religion, birthdate, employers, search activity and check-in locations.

Although the Internet gives convenient real-time information services to the people, the potential threats to confidentiality, integrity and availability (CIA) need to be addressed more effectively and permanently [4]. To fortify the security aspects of web-based servers and systems, Intrusion Detection Systems (IDSs) can be used as a complimentary device to many existing security appliances such as firewalls, password authentication, access control and vulnerability assessments.

An IDS is an application system or device that functions to identify either hostile activities or policy violation activities within a network. IDSs play an active role in network surveillance and also functions as a network security guard and have been widely used in recent years as a network security component. They are employed to capture and analyze traffic movement and send an alarm when intrusive actions are detected. The alarm alerts the security analyst, who then takes necessary action. In general, IDSs can be classified as either a host-based IDS (HIDS) or as a network-based IDS (NIDS) to recognize signs of intrusion. The classification is based on whether or not the position of the IDS is meant either to capture traffic just for a specific host or for the complete network. In NIDS, the IDS is normally installed before and after the firewall so that traffic for the whole network segment is captured. In the case of HIDS, the IDS focuses on a specific host to examine packets, logs and system calls. As such, HIDS are more suitable for identifying internal attacks compared to NIDS.

There are two main types of IDSs: signature-based and anomaly-based. Although a signature-based approach is

the most widely used in commercial IDSs, we cannot write a signature to detect unknown attacks beforehand. On the other hand, anomaly-based IDSs can detect unknown attacks, but they have problems that a low detection rate and a high false positive rate. To overcome these shortcoming, many researchers have been developing high performance anomaly-based IDSs.

Although Anomaly-based IDSs (AIDSs) can detect unknown attacks, they still have problems except for a detection performance. Since an AIDS just classifies network traffic into normal or abnormal, AIDS operators have to manually inspect an alert to identify whether an unknown attack exists or not. Moreover, AIDS reports a lot of alerts related to known attacks. It is very difficult to manually detect only unknown attacks from AIDS alerts.

II. ATTACK TYPES

The common categories of class attack as below:

Probe

A probe attack is an attempt to gather or learn specific information in a targeted network or host for reconnaissance reasons (e.g., port scanning). This type of attack is commonly used by an attacker to retrieve information from the machines connected inside the network where the host is vulnerable to this type of attack depending on the type of operating system or version of software installed or used. This kind of attack functions as a preliminary stage for an attacker before they launch an attack which purports to actually compromise the targeted network or host. This class of attack is the extremely common since it requires very little technical skill. Although there is no specific destruction to an organisation caused by these activities, they are still considered a serious threat due to the information obtained by the attacker, which is likely to be useful in launching any future attacks.

SQL Injection attack

SQL is a programming language used to communicate with databases. Many of the servers use SQL to store and manage the data in their databases. This type of attack specifically targets this kind of server, using malicious code to get the server to divulge information it normally wouldn't. Usually the server stores private information of the users such as credit card numbers, usernames, passwords or other personal information which may bring profit to the attacker. It works by exploiting any of the known SQL vulnerabilities that allow SQL server to run malicious code. For example, if a server is vulnerable to an injection attack, it may be possible for an attacker to go to a websites search box and type in a code that would dump all the stored data into attacker's database.

Cross-Site Scripting (XSS):

In this type of attack, an attacker targets a vulnerable website and target its stored data, such as user credential data or financial data. It is similar to SQL injection attack, where the attacker injects malicious code into a website. In this case, the website is not being hacked instead the malicious code runs only in the user's browser when they visit the attacked website. One of the ways to deploy this attack is by injecting malicious code into a comment or a script that can run automatically. It can significantly damage a website's reputation by placing the users' information at risk without any indication that anything malicious even occurred.

Denial of Services (DoS):

Denial of Services attacks are class of attack where an attacker sends a huge volume of request connections, normally with the intention of disrupting and paralysing the system server. In short, the attack encompasses destructive characteristics aimed at compromising the targeted network system infrastructure. One example of a DoS attack is when a web service is rendered unable to respond to legitimate users who need access because the server is flooded with innumerable connection requests. DoS attacks are classified based on the degree to which they cause unavailability of service to legitimate users.

User to Root (U2R):

The user to root attack is a type of attack during which an attacker exploits the administrative account to gain access to the root in an attempt to retrieve, modify or abuse important resources inside the system. Social engineering is a common method used as part of this attack, involving the attacker gaining access to the victim's account and exploiting a vulnerability in order to gain access as a super user. An example of this kind of attack is buffer overflow, where the attacks is the cause of regular programming errors or system settings mistake.

Remote to User/ Remote to Local (R2L):

Remote to user attacks are also known as remote to local attacks. This type of attack happens when an attacker exploits a vulnerability in the victim's machine over the network to illegally gain local access as an authorised user.

The privilege of this successful attack allows the attacker to gain the status of an authorised user to perform legitimate activities. These common attacks usually involve social engineering. Commonly, the attacker uses a trial-and-error approach by determining the user's password perhaps through some scripting method such as a brute force attack. Some sophisticated approaches involve the attacker successfully installing malicious tools with the intention of capturing the user password before using it to gain access to the system.

III. UNKNOWN ATTACK DETECTION METHODS

In [5], the Logitboost based algorithm was used to detect unknown attacks. Initially, the dataset was divided into training and testing dataset in 30% and 70% respectively. The dataset was preprocessed to select relevant features using hybrid feature selection technique, that is it combines both wrapper based and filter based methods. This feature selection technique is applied only to the training dataset. Then the Logitboost algorithm was then employed as a meta classifier together with random forest as a weak classifier. The proposed method was analysed through experiments using two different datasets: NSL-KDD dataset [6] and UNSW-NB15 [7] dataset. These datasets are publicly available online and has been used by many researches in this field. The NSL-KDD dataset is the traditional and most commonly used dataset in this field. In essence, the dataset is a modified version of the KDD Cup 1999 dataset, with some redundant traffic removed. In contrast, the UNSWNB15 dataset is a modern updated dataset, which claims to contain more realistic and modern attack types. The care was taken while dividing training and testing dataset that some attack instances were different in both the dataset.

The performance metrics used to analyse this method were false alarm rate (FAR), detection rate (DR) and accuracy (ACC). False alarm rate is the amount of benign traffic detected as malicious traffic. Detection rate is the proportion of detected attacks among all attack data. Accuracy is the number of instances correctly predicted in percentage form. The NSL-KDD dataset contains 41 features it was reduced to 10 features and UNSW-NB15 dataset contains 43 features it was reduced to 5 features after applying hybrid feature selection.

There are, in total, 9 types of attack in the NSL-KDD dataset. In the training dataset, there are 5 types of attack present: back, apache2, neptune, portsweep and saint whilst 8 types of attack: back, apache2, neptune, portsweep, ipsweep, satan, nmap and phf are in the testing dataset. It can be seen that 4 out of 8 types of attack "ipsweep, satan, nmap and phf" in the testing dataset are new attacks, which are not available in the training dataset. Among all the attack traffic present in the testing data, our proposed ensemble approach successfully recognised 99.10% instances of attack traffic. The attack type with the highest detection rate is DoS with 99.75%, followed by Probe with 54.83% and the lowest is the R2L with 16.67%. As a result of further investigation, the poor performance of R2L was determined since the connection of R2L and normal is similar, it is almost impossible for the system to distinguish between these two classes.

The same approach on the UNSW-NB15 dataset was applied. This dataset is comprised of real normal traffic combined with a variety of imbalanced synthetic attack traffic, which results in this dataset being more challenging to evaluate. In the training dataset, there are 7 types of attack present: backdoor, fuzzers, reconnaissance, exploits, dos, worms and generic whilst 8 types of attack: backdoor, fuzzers, reconnaissance, exploits, analysis, DoS, worms and generic are in the testing dataset. The main difference between the testing data and the training data is that it contains a new attack type named "analysis". The proposed ensemble approach successfully obtained an 89.75% detection rate among all attack traffic existing in the testing dataset.

The attack type with the highest detection rate is backdoor with 100% detection, followed by worms with 99.12%, reconnaissance with 98.75%, exploits with 94.33%, generic with 91%, dos with 87.5%, fuzzers with 80.98% and the lowest is analysis with 6.63%. The results show that with respect to five out of eight types of attack, our approach achieved a detection rate of more than 90%. The low detection rate of "analysis" is due to the unavailability of samples residing in the training dataset, which make it difficult for the system to classify it as an attack. Even after achieving the lowest detection rate, the system is still able to recognise "analysis" 6.63% of the time.

In addition, the performance of proposed approach was evaluated with some eminent state-of-the-art data mining algorithms used in IDSs such as: Naïve Bayes, Support Vector Machine, Multilayer Perceptron, Decision Tree, Random Forest and Adaboost. Five single classifiers are evaluated individually in terms of the time taken to build classification models, detection time, false alarm rate, detection rate and accuracy rate to choose a better combination for the Logitboost classifier.

Table -1: comparison of FAR, DR, ACC with other six algorithms in NSL-KDD dataset [5]

Algorithms	Detection Time (sec)	False Alarm Rate (%)	Detection Rate (%)	Accuracy (%)
Naïve Bayes (NB)	0.11	19.18	42.73	53.61
Support Vector Machine (SVM)	7.37	32.55	87.00	87.41
Multilayer Perceptron (MLP)	0.08	6.50	53.43	64.86
Random Forests (RF)	0.06	7.89	89.32	90.11
Decision Tree (J48)	0.04	6.68	88.23	89.68
Adaboost + Random Forest	0.74	8.30	89.71	90.27
Logitboost +Random Forests (RF)	0.65	8.22	89.75	90.33

In the NSL-KDD dataset, as shown in table 1, RF had shown comparable performance in terms of the accuracy, detection rate and false alarm rate. Although J48 had shown a faster detection time by 50% over RF, the detection and accuracy rate achieved by RF is slightly better than J48. Meanwhile, in the UNSW-NB15 dataset, RF outperformed every single other classifier by achieving 90.11% detection accuracy. So, this approach provides a comparable detection accuracy rate with a low false alarm rate.

In [8], anomaly detection method is used to detect the unknown attacks. The overall process of proposed method is composed of following 4 steps (Fig.1).

Step 1: Anomaly Detection- Detect attack traffic from Kyoto2006+ dataset.

Step 2: Feature Extraction- Extract 10 features from anomaly-based IDS alerts and make training data and testing data.

Step 3: Training- Applying one-class SVM to training data.

Step 4: Testing- Analyse testing data with one-class SVM Model and classify it into unknown or known attacks.

Kyoto2006+ dataset [9] was used to evaluate the proposed approach. It is obtained from a honeypot networks of Kyoto University. In the honeypot networks, several types of honeypots are deployed over 5 different networks which are outside and inside of Kyoto University. Kyoto2006+ dataset deploys a mail server in the same network to collect for normal traffic. From traffic data of the network, Kyoto 2006+ dataset extracts 14 conventional features and 10 additional features for each session. The former 14 features are extracted based on KDDCup 1999 dataset that is widely used for performance evaluation in intrusion detection system. The latter 10 features are extracted for more effective investigation. For example, signature-based IDS alerts, Antivirus alerts, source IP address and port number, time the session was started and so on.

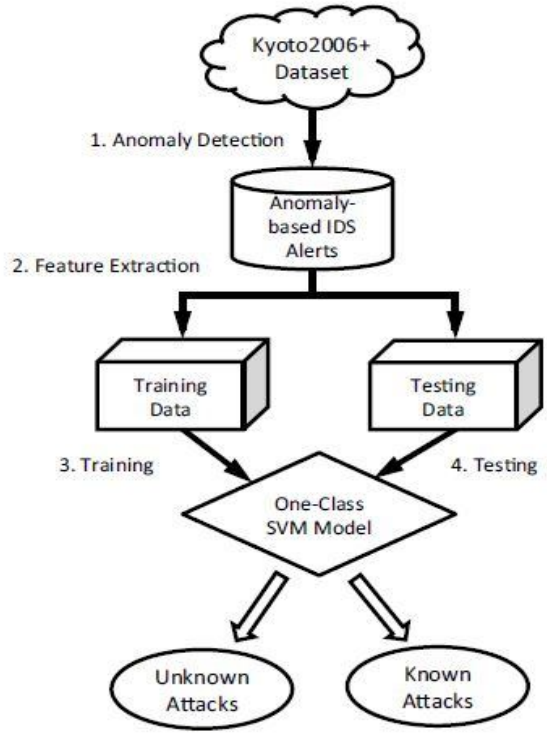

Fig-1: Steps of overall process [8]

The proposed approach of anomaly detection method is divided into two phases they are:

Training Phase
1) Filtering: filter out attack data from the training data.
2) Clustering: separate the filtered data into k clusters.
3) Modelling: apply one-class SVM to each k clusters.
Testing Phase
1) Dividing: assign the testing data to the closest cluster.
2) Classifying: classify the test data as normal or attack using corresponding one-class SVM model.

After anomaly detection, 10 features were extracted from AIDS alerts. Some assumptions were made on the characteristics of unknown attacks. 3 features are used directly in AIDS alerts; duration, source bytes and destination bytes. If unknown attacks trigger alerts, these features are irregular for each alert. Otherwise, these features have approximately constant values. The rest of features are extracted from AIDS alerts using a method proposed by Song [10]. They extract 7 statistical features. Because their method is specialized in feature extraction from signature-based IDS alerts, it cannot extract appropriate feature values from AIDS alerts. In Training Phase one-class SVM is applied to the above 10 features to detect unknown attacks from anomaly-based IDS alerts. One-class SVM seeks a hypersphere that includes most of training data within it. Because almost all of alerts is known attacks, inside the hypersphere can be considered known attacks, while outside is unknown attacks. In Testing Phase testing data is compared with SVM model. If a data instance is inside the hypersphere, the data is regarded as known attacks. Otherwise, it is regarded as unknown attacks. The detection rate is 80.0%. The approach detected 1,404 alerts as unknown attacks. Without this method, IDS operators have to manually analyze all AIDS alerts (34,436). Because this method significantly reduces the number of alerts which require manual analysis, it effectively supports their operational overhead. The proposed method has a higher detection performance against unknown attacks which do not trigger any signature-based IDS alerts.

IV. CONCLUSION

Many anomaly detection studies were conducted in the past. But, achieving exceptionally low false alarm rates with high attack recognition capabilities for unknown attacks is a major challenge. In this paper we have presented the two methods to detect the unknown attacks. The experimental results have demonstrated that the proposed approach has successfully recognised some unknown attacks and achieved comparable performance. Moving forward, the final successful results will be transformed into signatures and stored inside a database. By doing this, detection time can be drastically reduced, since the new entry traffic can be matched with benign/malicious signatures generated from previous detection.

REFERENCES

[1] W. Koff and P. Gustafson, “CSC LEADING EDGE FORUM Data rEvolution,” CSC LEADINGedgeforum. Tech. Rep. 68, 2011.
[2] S. Jones, “NHS seeks to recover from global cyberattack as security concerns resurface,” 2017. [Online]. Available: https://www.theguardian.com/society/2017/may/12/hospitals-across-england-hit-by-large-scale-cyber-attack. [Accessed: 02-Jun-2017].
[3] The New York Times, Facebook security breach exposes accounts of 50 million users. [Online] Available: https://www.nytimes.com/2018/09/28/technology/facebook-hack-data-breach.html
[4] S. V. Thakare and D. V. Gore, “Comparative Study of CIA,” 2014 Fourth Int. Conf. Commun. Syst. Netw. Technol., pp. 713–718, 2014.
[5] Muhammad Hilmi Kamarudin, Carsten Maple, Tim Watson, Nader Sohrabi Safa, ”A LogitBoost-Based Algorithm for detecting Known and Unknown Web Attacks” IEEE Access- vol.5, pp. 26190 – 26200, 2017.
[6] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” IEEE Symp. Comput. Intell. Secur. Def. Appl. CISDA 2009, no. Cisda, pp. 1–6, 2009.
[7] N. Moustafa and J. Slay, “The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set,” Inf. Secur. J. A Glob. Perspect., vol. 25, no. 1–3, pp. 18–31, 2016.
[8] Masaaki Sato, Hirofumi Yamaki, Hiroki Takakura, “Unknown Attacks Detection Using Feature Extraction from Anomaly-based IDS Alerts”.
[9] “Kyoto2006+dataset,” http://www.takakura.com/Kyoto data/.
[10] J. Song, H. Takakura, and Y. Kwon, “A generalized feature extraction scheme to detect 0-day attacks via ids alerts,” in Applications and the Internet, 2008. SAINT 2008. International Symposium on. IEEE, 2008, pp. 55–61.