ISSN:0975-9646

Autonomous Network Security using Unsupervised Detection of Network Attacks

Pragati H. Chandankhede
Dept.of Computer Science and Engg.
G.H.Raisoni College of Engg .
Nagpur, India

Sonali U. Nimbhorkar
Assistant Professor, Dept. of Computer Science
G.H.Raisoni College of Engg
Nagpur,India

***Abstract*—As a use of network increases for critical transaction, so huge damages are caused by intrusion attacks hence there is the need of Computer network security. To protect network against various active and passive attack, various technique have been suggested .Traditional methods depends on specialized signatures of previously seen attacks, or on expensive and difficult to produce labeled traffic datasets for profiling and training. In this paper, network attack are detected in unsupervised way without following the traditional way of signatures or labeled traffic.**

Keywords-component; Anomaly Detection; Intrusion Detection System; Automatic Generation of Signature; Autonomous Security

I. INTRODUCTION

Network attack detection is the very challenging task for the network operator in today's internet. It is being challenging task because network attack are moving targets , they are not steady. Attacker may launch every time new attack, which is not seen previously . So there is the need of detection system that will be able to detect various attacks of different range and with variety of characteristic[3]. This detection system should use the very less amount of previous knowledge or no use of any type of information at all.

In research literature and commercial detection systems there are two different approaches namely signature based detection and anomaly detection for detection of attack. Signature based detection relies on the use of specifically known pattern of unauthorized behavior. It depends on sniffing packet. It monitors and compares the packet with predetermined attack patterns which are also known as signature. That is signature based detection system is used to detect those attacks which they are program to alert on. This detection system cannot defend against unknown attack. On the contrary, Anomaly detection builds normal operation traffic profile which detects anomalies as the activities that deviate from the baseline[5]. Thus it uncover abnormal pattern of behavior. This detection system can detect new, previously unseen attack. But as it has to build normal operation profile and it require training to construct normal operation profile and hence it is being time consuming. The new services and applications are constantly emerging and this type of detection is being difficult.
In this paper the approach is used to detect both known as well as unknown attack . this is done by the production of signature that determine the attack in an online basis.algorithm that is being apllied for characterizing attack will run in consecutive three stages, which is being represented by flow as shown below in figure 1.

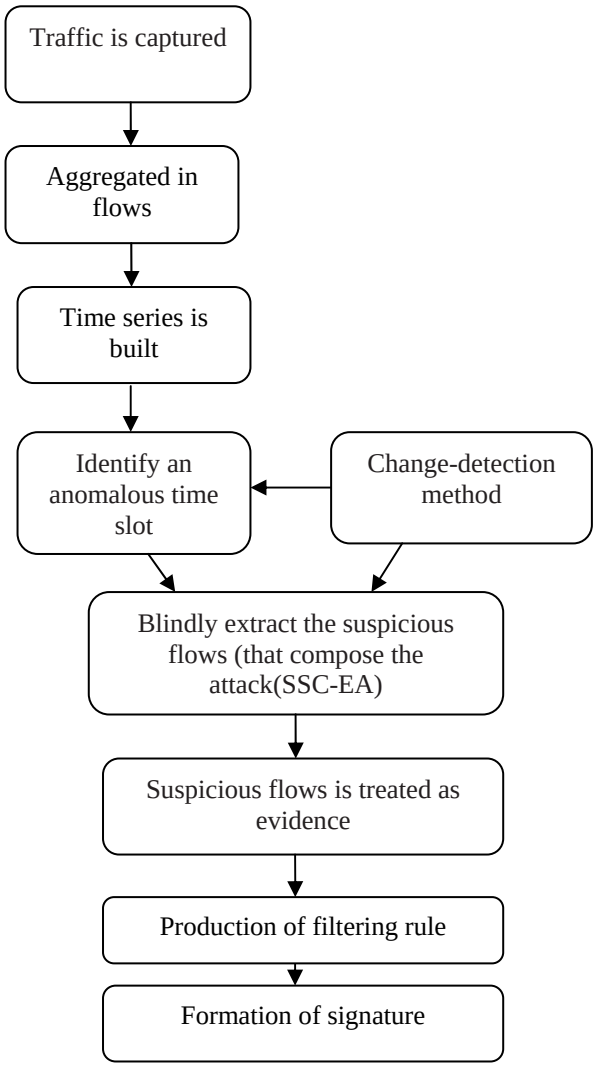

Figure 1: Flow of algorithm

The algorithm follows in three consecutive stages. Firstly, using a temporal sliding-window approach, traffic is captured and it is aggregated in flows. This is done using different levels of traffic aggregation. For simple traffic metrics such as number of bytes, flows in each time slot, time series are built. And any change-detection method is applied to identify an anomalous time slot.

In the second stage unsupervised detection algorithm begins. It uses the output of first stage as the input. Subspace clustering(SSC) and Evidence Accumulation (EA) provides method which can extract the abnormal or suspicious flow from the previous output.SSC and EA will provide the traffic structure which further can be used to produce filtering rule. Filtering rule helps to provide characteristic of attack[8] .But when the network operator deal with the unknown attack, the characterization of attack may became much more difficult as it require good, easy, simple information as the input .To remove this issue, new traffic signature is developed by combining relevant filtering rules. This signature will detect the attack coming in future; this is the important step toward autonomous security.

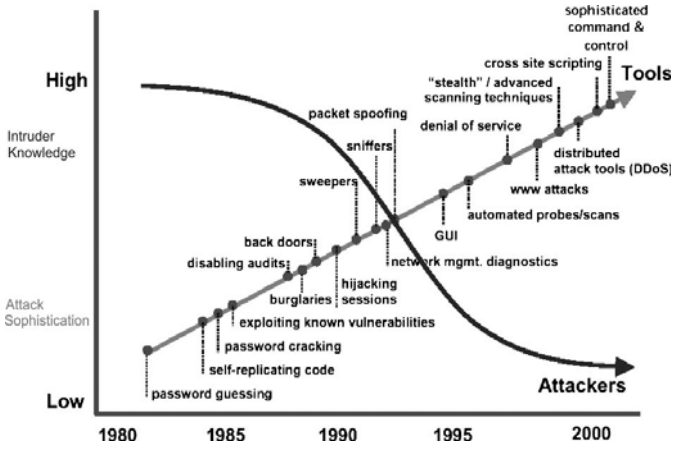

Figure 2: Attack sophistication vs. intruder technical knowledge

II. PREVIOUS WORK

Network attack is any activity that compromises the stability and security of information. It include the activity such as destabilizing the network as the whole, gaining unauthorized access to the file or some time it is simply mishandling the software. The main objective of building the autonomous system is to detect such intrusion attack by scanning automatically the network activity. Because of the commercialization of the Internet, intrusion incidents to computer systems are increasing. Due to its extended network connectivity, Computer systems are turning out to be more and more susceptible to attack. As it is not theoretically possible to set up a system with no vulnerabilities, intrusion detection has emerged as a significant field of research, from a large quantity of routine communication activities. Several machine learning, Intrusion detection has emerged as a significant field of research. to detect intrusion activities
(ML) algorithms, for instance Neural Network , Support Vector Machine , Genetic Algorithm ,Fuzzy Logic , and Data Mining have been extensively employed [6].

Over the years, attacks have become both increasingly numerous and sophisticated. Not only has there is a markable increase in the number of attacks, but along with that the sophistication and complexity has also increased. Thus many attacks are now relatively ‘‘user-friendly’’ and in-depth technical knowledge is no longer required to launch an attack. This has led to the rise of various groups of attackers, such as ‘‘script-kiddies’’, who while ignorant of how their attack works, can cause great damage. In Lipson (2002), this trend is represented graphically as shown in Figure 2. of Attack sophistication vs. intruder technical knowledge. Due to such scenario of network attack, it became important for the researchers and operators to know about trends in network traffic.

In the last few decades, Signature based detection technique is used for network attack detection in which; using network data the feature are extracted. These feature value are being compared with attack signature provided by human expert. Thus signature based detection relies heavily on use of specifically known pattern of unauthorized behavior.

Other approaches make the use of machine learning and mining of data techniques to train labeled network data. **Misuse Detection and Anomaly Detection** are the two major categories of data mining-based intrusion detection.

In anomaly detection system, system tries to define what is normal and then detect how analyzed data is different from normal model [8]. But meanwhile if some intrusion arises, it will not be considered as normal. It detects them initially. It is also possible that training data will contain traces of intrusion, so in such case future instance of the attack may not be detected rather, they will be treated as normal.

In misuse detection, set of labeled data is use to train the machine learning algorithm and the detection model is built. This detection model will be similar to the signature describe earlier. But this is also similarly vulnerable to new type of attack as the signature based method.

Various approaches are used to address the problem in anomaly detection. These approaches are dependent on analysis of data that is being available. Network data can be obtained at multiple levels of granularity such as end-user-based or network-based.

Today anomaly detection system should be able to detect a wide range of anomalies with diverse structure, using the minimum amount of previous information, or no knowledge at all.

It is being called as unsupervised anomaly detection. For that there is the use of KDD CUP data set[14]. For evaluating IDS, Lincoln Laboratory along with DARPA, launched DARPA 1998 dataset. It consists of testing data of two weeks and seven weeks of training data. The refined version of DARPA dataset that consist of only the network data is the KDD dataset. It consists of 4,900,000 connection vectors. Large number of unsupervised detection schemes proposed in

Literature survey is based on clustering and outlier detection technique.

Autonomous Network Security using Unsupervised Detection of Network Attacks will work in the different way. It is advantageous as the name suggests it is completely autonomous that is without any kind of calibration or previous knowledge, if it is plugged in monitoring system, it starts to work .Second advantage is that signatures build by the system are compact and easy which can characterize attack in effective way. Third and most important advantage is that it combines the robust clustering techniques such that many clustering problem are avoided.

III. UNSUPERVISED NETWORK ATTACK DETECTION

There are two knowledge based approaches, signature-based detection and anomaly detection as discussed above . IDSs, IPSs, and firewalls uses signature based detection. Signature based detection system can detect those attack which it is train to alert on. while anomaly detection uses labeled data for the creation of normal operation traffic profiles. But as this approach requires training for profiling, thus it becomes time consuming work. Thus this paper concentrates on tackling anomaly detection problem.there is the requirement of analysis technique which is not depending on knowledge, thai is knowledge independent technique.

Aiming at discovering Knowledge Independent system, new proposed algorithm is unsupervised network attack detection algorithm. The figure describing algorithm is as shown in figure 3.

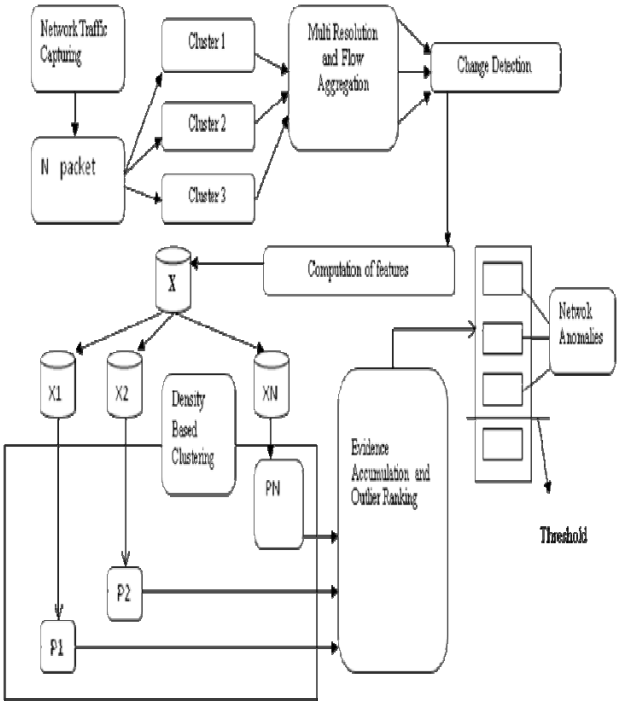

Figure 3: Description of unsupervised network attack detection algorithm.

Initially traffic is captured and packet are analyzed by aggregating them in multi resolution flow. On the top of these flow, different time series is built. And anomalous change is defined by change-detection algorithm based on time-series analysis.

2. Determining degree of abnormality:

There is the use of robust clustering algorithm like Sub-Space Clustering (SSC) , Density-based Clustering , and Evidence Accumulation Clustering (EAC) as combination of these approaches for providing traffic structure. These traffic structure are used as the evidence for determining by how much degree the traffic is not normal.thus the output of second stage are outlying flow.

3. Declaring anomalies:

Using a simple threshold detection approach, outlying flow which are top ranked are flagged as anomalies.

IV. MULTI-RESOLUTION FLOW AGGREGATION AND CHANGE DETECTION

Unsupervised network attack detection algorithm performs unsupervised anomaly detection. These are captured in consecutive time slot of fixed length. These are further aggregated in IP flows. At different flow-resolution levels IP flow are additionally aggregated using 9 different aggregations key. Thus there is coarser to finer-grained resolution.

Now to detect anomalous time slot, time series is built for traffic metrics which include IP flows per time slot, number of bytes, packets. For doing this aggregation key is used. Change detection method is then used on time series, such that at arrival of every new time slot, change detection method analyses different time series by using each aggregation key.

V. UNSUPERVISED ATTACK DETECTION THROUGH CLUSTERING

IP flows in the flagged time slot are used as the input for unsupervised attack detection. At this step unsupervised network attack detection algorithm ranks the degree of abnormality of every flow by using using clustering and outliers analysis techniques. For doing so, at two different resolutions, using either IPsource or IPdestination aggregation key IP flows are analyzed. There are two different anomalies on the basis of which traffic anomalies can be classified, 1-to-N anomalies and N-to-1 anomalies. When many IP flows are transferred from same source to different destination they are said to be 1-to-N anomalies. Example of 1 to N includes worms or virus. Likewise N-to1 means IP flows when transferred from different sources to one destination. Examples consist of DDoS attacks and flash crowds. 1-to-N anomalies are highlighted by IPsource. While N-to-1 anomalies are more easily detected with IP dst key. Even there are highly distributed anomalies, but the use of both key i.e, IPdestination key and IPsource key number of IP flows, can be

represented as outliers. Unsupervised network attack detection algorithm is based on clustering technique. Homogeneous groups of similar characteristics or clusters are formed by partitioning a set of unlabeled samples. Outliers are those Samples that do not belong to any of these clusters. It is important to identify the cluster properly to determine the outlier. Our aim is to determine or ranking how much different these are. Different partitions of data are produce by Different clustering algorithms. Or different results are produce, even the same clustering algorithm are used by using different initialization parameters. Thus present clustering algorithm are not robust. To remove this major drawback of lack of robustness. This is done by using the notions of clustering ensemble and multiple clusterings combination. For unsupervised network attack detection the combination of Subspace and evidence accumulation clustering is used.

VI. CONCLUSION

The completely unsupervised algorithm for detection of network attacks has many interesting advantages with respect to previous proposals. Exclusively unlabeled data is used for detection and characterization of network attacks, it does not depend or assume any signature, model, or data distribution. Thus new previously unseen attack can be detected, without using statistical learning. Robustness is removed by combining the notions of Sub-Space Clustering and multiple Evidence Accumulation, the algorithm avoids the lack of robustness of general clustering approaches, improving the power of discrimination between normal-operation and anomalous traffic.

VII. REFERENCES

1.E. Kohler, J. Li, V. Paxson, and S. Shenker. Observed Structure of Addresse in IP Traffic. In Internet Measurement Workshop, Marseille, November 2002

2.Y. Zhang, S. Singh, S. Sen, N. Duffield, and C. Lund. Online Identification of Hierarchical Heavy Hitters: Algorithms, Evaluation, and Applications. In Internet Measurement Conference, Taormina, Italy, October 2004

3. S. Kim and A. L. N. Reddy. A Study of Analyzing Network Traffic as Images in Real-Time. In IEEE International conference of communication, 2005.

4. S. Kim, A. L. N. Reddy, and M. Vannucci. Detecting Traffic Anomalies through Aggregate Analysis of Packet Header Data. In Networking, 2004.

5.A. Lakhina, M. Crovella, and C. Diot. Characterization of Network-Wide Anomalies in Traffic Flows (Short Paper). In Internet Measurement Conference, 2004.

6.A. Lakhina, M. Crovella, and C. Diot. Diagnosing Network-Wide Traffic Anomalies. In ACM Special Interest group on Data Communication, Portland, August 2004.

7.K. Xu, Z.-L. Zhang, and S. Bhattacharyya. Profiling Internet Backbone Traffic: Behavior Models and Applications. In ACM Special Interest group on Data Communication, 2005.

8.Pedro Casas,Johan Mazel .Steps Toward Autonomous Network Security :Unsuperised Detection of Network Attacks,IEEE International conference of communication , 2011

9. Wiilliam Stalling Cryptography and Network Security , third edition.

10. D. Fasulo, \An analysis of recent work on clustering," Tech. Rep., University of Washington, Seatle. http://www.cs.washington.edu/homes/dfasulo/clustering.ps. http://citeseer.nj.nec.com/fasulo99analysi.html, 1999.

11. D. Judd, P. Mckinley, and A. K. Jain, \Large-scale parallel data clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 2, pp. 153{158, 1997.

12.S. K. Bhatia and J. S. Deogun, \Conceptual clustering in information retrieval," IEEE Trans. Systems, Man, and Cybernetics, vol. 28, no. 3, pp. 427{536, 1998.

13 C. Carpineto and G. Romano, \A lattice conceptual clustering system and its application to browsing retrieval," MachineLearning, vol. 24, no. 2, pp. 95{122, 1996.

14. L. Parsons et al., “Subspace Clustering for High Dimensional Data: a Review”, in ACM SIGKDD Expl. Newsletter, vol. 6 (1), pp. 90-105, 2004.