Detection of Unknown Anomaly on Network - The Emerging Way

Anup G. Kadu#, Dr. A.S.Alvi*

#*M.E., Information Technology,P.R.M.I.T.& R,Badnera* - **Information Technology,P.R.M.I.T.& R,Badnera*

1anupkadu@gmail.com
2abraralvi@gmail.com

Abstract--- **The existing knowledge-based approaches are not sufficient to solve the anomaly detection problem, and to that a good solution should also include knowledge-independent analysis techniques. There are some algorithms, and it becomes critical in the case of unsupervised detection, because there is no additional information to select the most relevant set some approaches can be easily extended to detect other types of attacks which is unknown to the system, considering different sets of traffic features which depend on the how we create log file. In fact, more features can be added to any standard list to improve detection and characterization results, which means more the feature more will be the correctness. The Detection of Unknown Attacks on Network is simply to detect the attacks which are completely unknown to us. There is no prior knowledge about that data. There are some knowledge based algorithms in existence which are used for network security but they are inefficient as they are (Signature Based and Anomaly Based) whenever there is a vast amount of continuous incoming data then it is a big risk regarding the network attacks which are knowledge based. Our particular goal is to detect those network attacks with the help of Robust Clustering Algorithm and make complete system secure.**

Keywords-- **Signature Based, Anomaly Based, Robust Clustering**

I. INTRODUCTION

The unsupervised detection of network attacks represents an extremely challenging goal of detecting attacks in today's ever growing traffic without any prior knowledge**.** The detection of network attacks is a high priority task for network operators in present network traffic. Denial of Service attacks (DoS), Distributed DoS (DDoS), malware, Trojan, network/host scans, and spreading worms or viruses are examples of the different attacks that daily harm the integrity and normal operation of the network. The main mission is that we have to detect those harmful attacks in a ever growing network traffic in which target are moving.

As we said earlier their two knowledge based approaches we found in literature review are very effective and working well those approaches are: signature-based detection and anomaly detection. Signature-based detection systems are highly effective to detect those attacks which are already known by the system that means it detect only those attacks for which signature is already created. And if the attack come which is already not known by the system then system is not able to detect those attacks, if some want to detect those unknown attacks then he/she want to create the signature for this unknown attacks first then they become able to detect those attacks. But problem is that, in this case the creation of signature requires much time and it's also very cost effective process which causes delay in the detection of network attacks. The another approach that we found in literature is also a good approach called Anomaly detection this approach is able to detect those attacks which is completely unknown to the system for it uses labelled traffic to create normal operational traffic profile and mark this normal operational traffic profile as baseline and then after this any traffic that deviate from this baseline it treat as anomaly this is a good approach of detecting unknown attacks but the problem is that it require normal operation traffic profile, and to create this it require training which is very time being process and also it is very difficult to maintain the normal operation traffic profile. We also found that in literature Signature engines also have their disadvantages. Because they only detect known attacks, a signature must be created for every attack, and novel attacks cannot be detected. Signature engines are also prone to false positives since they are commonly based on regular expressions and string matching. Both of these mechanisms merely look for strings within packets transmitting over the wire. A disadvantage of anomaly-detection engines is the difficultly of defining rules. Moreover detailed knowledge of normal network behaviour must be constructed and transferred into the engine memory for detection to occur correctly. On the other hand, once a protocol has been built and a behaviour defined, the engine can scale more quickly and easily than the signature-based model because a new signature does not have to be created for every attack and potential variant.[1] [2]

Our approach relies on robust clustering algorithms to detect both well-known as well as completely unknown attacks, and to automatically produce easy-to-interpret signatures. The analysis is performed on packet-level traffic, captured in consecutive time slots of fixed length *t* and aggregated in IP flows standard *for one second*. IP flows are additionally aggregated at 9 different *flow* levels l_i. These include: *source IPs*, *destination IPs*, *source Network Prefixes*, *destination Network Prefixes*, and *traffic per Time Slot*. The complete anomaly detection algorithm runs in three successive steps. The first step consists in detecting an anomalous time slot where an attack might be hidden. The unsupervised anomaly detection algorithm start in the second step, using as input the set of IP flows. The method uses robust clustering techniques based on Sub-Space Clustering and Evidence Accumulation to extract the suspicious flows that compose the attack and rank those flows according to their abnormality. In the third and final step, we mark the anomaly according to their abnormality. The characterization of an attack can be a hard and Time-consuming task, especially when dealing with unknown attacks. Experienced operators can be quickly mentally tired if simple and easy-to-interpret information is not provided to prioritize the time spent in the analysis. Now some rules are combined in to the traffic features to solve this problem from which one can able to create signature. This signature can ultimately be integrated to any standard security device to detect the attack in the future, which constitutes a major step towards automatic network security that means algorithm automatically produces new signatures without any previous data about traffic or knowledge about the attack. [3]

II. RELATED WORK AND BACKGROUND

Two different approaches are by far dominant in current research literature and commercial detection systems: signature-based detection and anomaly detection. When an attack is identified, normally after its occurrence during a diagnosis phase, the associated anomalous traffic flow pattern is coded as a signature by human experts, which is then used to detect occurrence of the same attack in future. Signature-based detection methods are highly effective to detect those attacks which is already known by the system that means those attacks for which signature is already created. However, they cannot prevent from the new attacks, simply because they cannot recognize what they do not know.

And for building new signatures it require much time and it’s also a time consuming task. On the other hand, anomaly detection uses labelled data to build normal- operation-traffic profiles, detecting anomalies as activities that deviate from this baseline. Such methods can detect new kinds of network attacks not seen before. And in this case it requires normal traffic to create normal operation traffic flow which is very difficult to obtain it also need training to create normal operation traffic profile. Labelling traffic as anomaly-free is very time consuming and expensive process. It is difficult to guarantee that no anomalies are buried inside the collected data. In addition, it is not easy to keep an accurate and up-to-date normal-operation profile. Our objective is that these two knowledge-based approaches are not sufficient to solve the anomaly detection problem, and to this a good solution includes unsupervised detection of network attack. To this aim we propose UNADA, an Unsupervised Network Anomaly Detection Algorithm that detects network traffic anomalies without relying on signatures, training, or labelled traffic of any kind. UNADA relies on robust clustering algorithms to detect outlying traffic flows.[4] [5]

The proposed system will be identified to provide a solution to the problem of anomaly detection which is completely Knowledge Independent. In this we evaluate the ability of UNADA to discover network attacks in real traffic without relying on signatures, learning, or labelled traffic.

III. SYSTEM DESIGN

In the system design input data at first that contain the data packets over network traffic. A data packet in network traffic is an ordered sequence of object, this may contain anomaly and we have to detect those anomalies in the data packets over the network traffic. To detect those anomalies in the huge dataset we have to apply robust clustering approach which will create automatic signature. In my proposed work I am going to implement completely blind approach so for that no any previous knowledge about the anomaly and to detect such types of blind attack I am going to apply robust clustering approach for the detection of network anomaly in an completely unsupervised fashion .

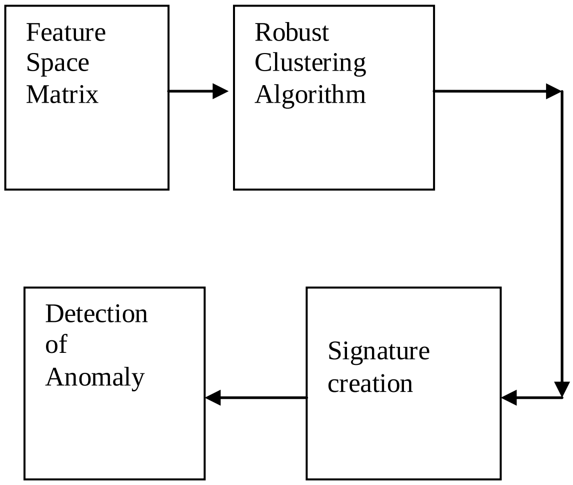


Fig.1 Organization of the system

Work plan define the specific steps as follows:

- Create the log file with the help of software or hardware.
- Separate the data according the feature of traffic captured in log file.
- Apply sliding time windowing scheme for every 1sec.
- Aggregate flow of traffic for specific time slot.
- Creation of feature space matrix by using following similarly we have to create feature space matrices for all time slot.
  i.e., X= € (x1, x2…….xn)
- Apply clustering algorithm and create cluster according the requirement.
- Detect outlier using outlier detection algorithm.
- Declare smallest group of cluster as outlier
- Trace back outlier in feature space matrix, aggregation and log file.
- Trace data to Create signature for anomalous flow.
- Logged the signature in the signature table and update the signature table.

A. Advantages the proposed system on the existing Approaches:

This algorithm has some advantage over the existing approach.

1) It works in a completely knowledge Independent manner, which means that it can be directly embed with any network monitoring system, without any kind of previous knowledge.

2) It uses robust clustering techniques to avoid general clustering problems such as sensitivity to initialization, specification of number of clusters.

3) It automatically builds compact and easy-to-interpret signatures to detect network attacks, which can be directly integrated into any traditional security device.

4) It is designed to work on-line, for that we have just embed it on hardware and execute it for the detection network attack, it will work in an online basis.

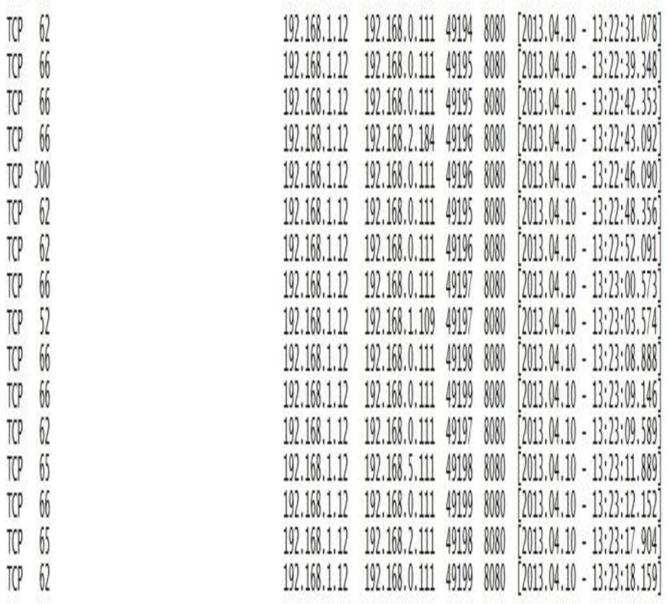
IV. TESTING

For testing we first create many log file in different network and then we run our algorithm for these entire log file repeatedly & we get the proper result. We also add signature in the signature table and then next time we check the same log file for anomaly but this time we compare it with signature in the signature table and we get the same result.

To this we show one small log file, which has one anomaly that we detect manually very easily. And our project exactly detect that anomaly from this we can say that it work correctly with any data in any network.

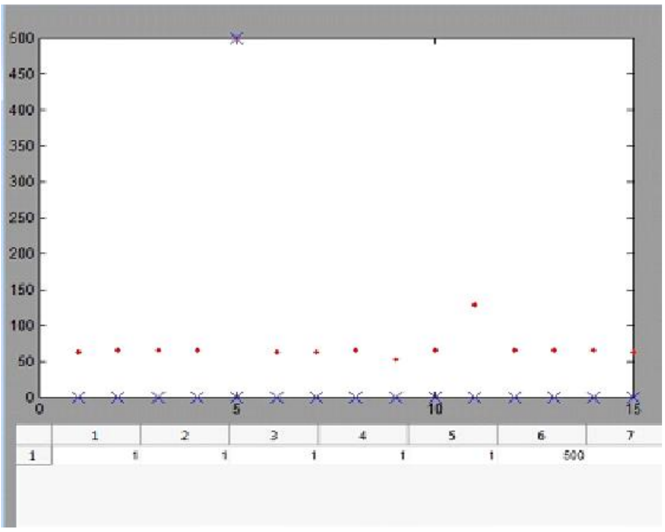
Log file containing anomaly in 5th row.

Table 1: Log File



```
TCP  62     192.168.1.12  192.168.0.111  49194  8080  [2013.04.10 - 13:22:31.078]
TCP  66     192.168.1.12  192.168.0.111  49195  8080  [2013.04.10 - 13:22:39.348]
TCP  66     192.168.1.12  192.168.0.111  49195  8080  [2013.04.10 - 13:22:42.353]
TCP  66     192.168.1.12  192.168.2.184  49196  8080  [2013.04.10 - 13:22:43.092]
TCP  500    192.168.1.12  192.168.0.111  49196  8080  [2013.04.10 - 13:22:46.090]
TCP  62     192.168.1.12  192.168.0.111  49195  8080  [2013.04.10 - 13:22:48.356]
TCP  62     192.168.1.12  192.168.0.111  49196  8080  [2013.04.10 - 13:22:52.091]
TCP  66     192.168.1.12  192.168.0.111  49197  8080  [2013.04.10 - 13:23:00.573]
TCP  52     192.168.1.12  192.168.1.109  49197  8080  [2013.04.10 - 13:23:03.574]
TCP  66     192.168.1.12  192.168.0.111  49198  8080  [2013.04.10 - 13:23:08.888]
TCP  66     192.168.1.12  192.168.0.111  49199  8080  [2013.04.10 - 13:23:09.146]
TCP  62     192.168.1.12  192.168.0.111  49197  8080  [2013.04.10 - 13:23:09.589]
TCP  65     192.168.1.12  192.168.5.111  49198  8080  [2013.04.10 - 13:23:11.889]
TCP  66     192.168.1.12  192.168.0.111  49199  8080  [2013.04.10 - 13:23:12.152]
TCP  65     192.168.1.12  192.168.2.111  49198  8080  [2013.04.10 - 13:23:17.904]
TCP  62     192.168.1.12  192.168.0.111  49199  8080  [2013.04.10 - 13:23:18.159]
```

The following graph shows the result of unsupervised anomaly detection



	1	2	3	4	5	6	7
1	1	1	1	1	1	500	

IV CONCLUSION:

This is completely unsupervised approach for the detection of various unknown attack on network, since it detect the network attack without relying on previous signature, labelled data or any training.

It provides good security to the system from the unknown attack on network as it is able to detect unknown attack.

Additionally we have shown detection result that outperforms traditional approach for anomaly detection.

REFERENCES

[1] S. Hansman, R. Hunt “A Taxonomy of Network and Computer Attacks”, in *Computers and Security*, vol. 24 (1), pp. 31-43, 2005.

[2] Paul Barford, Jeffery Kline, David Plonka and Amos Ron “A Signal Analysis of Network Traffic Anomalies” In procedeengs of ACM Sigcomminternet Measurment Workshop 2002

[3] A. Soule et al., “Combining Filtering and Statistical Methods for Anomaly Detec-tion”, in Proc. ACM IMC, 2005

[4] Balachander Krishnamurthy, Subhabrata Sen, Yin Zhang Yan Chen_ AT&T Labs–Research; “Sketchbased Change Detection: Methods, Evaluation, and Applications”
180 Park Avenue University of California *IMC’03,* October 27–29, 2003.

[5] Ana L.N. Fred Telecommunications Institute Instituto Superior T´ecnico, Portugal and Anil K. Jain Dept. of Computer Science and Engineering Michigan State University, USA “Data Clustering Using Evidence Accumulation”2009.