Anomaly Extraction in Networks

Mr. Naushad Mujawar, Mr. Sohan Patil, Mr. Amit Kanase, Mr. Ravindra Jagadale, Prof. Gajanan Arsalwad.

(Information Technology, University of Pune, India)
(Information Technology, University of Pune, India)
(Information Technology, University of Pune, India)
(Information Technology, University of Pune, India)
(Dept. of Information Technology, Trinity College of Engineering & Research, Pune)

ABSTRACT : The application detects anomaly in network using techniques like histogram, cloning voting, filtering. To extract anomalous flows, one could build a model describing normal flow characteristics and use the model to identify deviating flows. We can compare flows of packets on network with previous flows, like new flows that were not previously observed or flows with significant increase/decrease in their volume. Identify an anomalous flow that combines and consolidates information from multiple histogram-based anomaly detectors [1] [4] [8]. Compared to other possible approaches. Build a histogram based detector that (i) applies histogram cloning[1][4], i.e., maintains multiple randomized histograms to obtain additional views of network traffic[3]; and (ii) uses the Kullback-Leibler (KL) distance to detect anomalies.

Keywords - Anomaly extraction, Apriori Algorithm, Association rules, Flow pre-filtering, Histogram cloning [1], voting.

I. INTRODUCTION

Anomaly in the network slows down the server, so that we can extract anomaly by using association rule, apriori algorithm and some techniques like histogram, cloning, voting, and filtering. Histogram shows traffic on the network with source and destination IP, packet size and number of packets which helps network administrator to take appropriate action [3]. Server generate alert message which tell us that anomaly is detect nor not.

A. Overview

An overview of our approach is to the Anomaly Extraction in the Network [2] describes the anomaly detection in the network with the help of Histogram technique, Cloning, Voting, Filtering, and by applying association rule; these whole things happen in the system, we considered as a server. Server analyze the traffic of the packets on the network, if the program finds something abnormal transferring of packets or packets beyond traffic limit which cause network failure, then server generate report on this situation which contains source machine address and/or destination machine address, no. of packets, time of sending. This report is for the Network Administrator to take appropriate action on the situation. Our project contain following approaches:

- **Histogram:** Use to represent the packet traffic on the network. It represent in the bar graph manner with time and no. of packets as a parameters.
- **Voting:** Voting is mainly used to extract useful information from raw metadata. Voting apply on the data generated by the histogram [3].
- **Flow Pre-Filtering:** Filtering is works as its name. It filters the suspicious data from large set of data.
- **Association Rule:** The filtered data generated by the Flow Pre-Filtering mechanism is converted into summary report by applying Association Rule on it.
- **Apriori Algorithm:** Apriori algorithm is the best data mining algorithm, which extract frequent item sets for the generation of data report.

B. Methodology

A server draws histogram (fig.1) of packet sent by the client with source IP, destination IP, source and destination port number, packet size and number of packets.

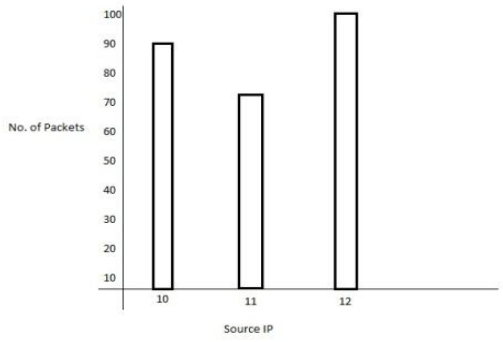

Fig.1 Histogram which shows no. of packets and Source IP.

After drawing histogram, server generate alert message. Depending on that message the network administrator will take action.

II. SYSTEM ARCHITECTURE

Application is designed for the LAN network, where it analyzes the network traffic and with the help of these data it can detect anomaly. In the proposed application the client sent the packet to the receiver via the system which acts as a router, (fig. 3) which analyses the IP addresses of sender and receiver. It also checks the network traffic with the help of histogram technique [3]. By voting technique we calculate the difference between normal data and suspicious dataflow. Then we filter the suspicious data. And by applying the association rule we detect the anomaly in the network.

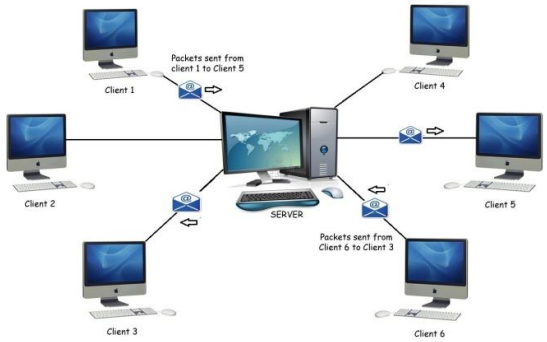

Fig.2 System Architecture.

III. EXISTING AND CURRENT APPLICATION

- **Existing System:** Identifying network anomalies is critical for the timely mitigation of events, like attacks or failures that can affect the security and performance of network. Traditional approaches to anomaly detection use attack signatures built in an Intrusion Detection System (IDS) that can identify attacks with known patterns. Significant research efforts have focused on building IDS's and, therefore, related production systems are presently employed in many networks. Although signature-based detection finds most known attacks, it fails to identify new attacks and other problems that have not appeared before and do not have known signatures.

- **Proposed System:** Our system contains four different modules. One is histogram representation that will observe the network traffic and alert the system upon anomaly detection [3]. Second module consists of histogram cloning which assures that the anomaly detected and finds the suspicious flows from network traffic. Third module is filtering which sort out the required data from the huge data. This sort out data is used to generate report about anomaly. Finally fourth module is to apply association rule to find the frequent item sets.

IV. REQUIREMENTS

Hardware Requirement

Processor: PIV – 500 MHz to 3.0 GHz

RAM: 512MB

Hard Disk: 10GB

Monitor: Any color display

Software Requirement

Operating System: Windows XP / Windows Vista / Windows 7 / Windows 8.

Development End: Java

IDE: Eclipse

V. ACTUAL IMPLEMENTATION

Finding an actual cause of anomaly

We are using a tool for finding an anomaly in the network is working as fast as previous tools are used for finding such like. We get an IP address of that machine which is mostly affect in the network

and eliminate the speed of network. We should also stop the service from that client which creates the anomaly situation.

Network like a LAN, MAN, WAN we are used for our need of today's life. In that network we have to provide a security for a good performance the result proper communication between entire networks.

Identifying a particular client that create such situation using a methods like histogram, voting, cloning, association rule that is very impressive for finding the hacker. We also block the packets coming from that particular pc which making a critical situation, so this is the biggest advantage of our tool that was not introduced in previous tool.

Fig.3 Client Interface for choosing destination to send packets.

Fig.4 Server Interface to start server.

VI. MODULE DESCRIPTION

Module 1: Server

Scope of Module or Component: Purpose of the Admin is to Login in the system and takes action on the basis of the report generated by application.

User Interface Design Model: Admin can Login into application through the login page.

Structure: Only one admin (network administrator) is requiring.

Function: Admin take action on the anomaly found on network on the basis of report generated by application.

Input & Output: Admin must Login into application and take control of the application.

Module 2: Histogram

Scope of Module or Component: Histogram describes the traffic on the network.

User Interface Design Model: Application designs the histogram for every system which transmits the packets.

Structure: Represents the network traffic in bar graph manner.

Function: Application receives the packets from sender and on basis of that the application draws the histogram for analyzing the traffic.

Input & Output: Amount of packets and the IP address are the input to the histogram; output is in the form of bar graph.

Module 3: Voting

Scope of Module or Component: Voting for select the l values out of k clones.

User Interface Design Model: Application conducts voting to select l values out of k.

Structure: Application select the few number of clones which contain suspicious traffic.

Function: Voting perform the operation of selection of clones which represents suspicious.

Input & Output: Traffic out of number of clones.

Metadata is the output.

Module 4: Filtering

Scope of Module or Component: Purpose of filtering module is to sort out the suspicious data from huge data.

User Interface Design Model: Flow Pre-Filter module is depends upon output of voting.

Structure: Output of voting is given to the filter module, where expected data can be sort out.

Function: Union of metadata is use to filter the suspicious data to obtain normal data.

Input & Output: Output of the voting module is input to the filtering module. Output is the suspicious flow on which the association rule is applying.

VII. ALGORITHM

Apriori algorithm: this algorithm is used for frequent item sets, these frequent item sets is used for constructing candidate item sets.

Ck: Candidate item set of size k

Lk : frequent item set of size k

L1 = {frequent items};

For (k = 1; Lk !=Ø; k++)

Do begin

Ck+1 = candidates generated from Lk;

For each transaction t in database do

Increment the count of all candidates in Ck+1 that are contained in t

Lk+1 = candidates in Ck+1

End

Return Uk Lk

VIII. CONCLUSION

Studied the problem of anomaly extraction. We present histogram based detector [4] [7]. We showed that rule mining is very effective. Newly occurred attacks can be detected. Anyone can send the number of packets; it is too difficult to find which one is unwanted. The existing one is having some disadvantages which we reduce by using better algorithm, formulas and technique.

IX. FUTURE SCOPE

Defense system: It can be used in defense system where server may be slow down by anomaly.

Aeronautical system: application used in time critical system.

Medical system: Application used in medical field where server wants to work faster.

Time critical system: Application works in any field where network require fast and server.

REFERENCES

Journal Papers:

[1] D. Brauckhoff, X. Dimitropoulos, A. Wagner, and K. Salamatian, "Anomaly extraction in backbone networks using association rules," in *IMC'09*, November 2009.

[2] D. Brauckhoff, M. May, and K. Salamatian, "Applying PCA for Traffic Anomaly Detection: Problems and Solutions," in *IEEE INFOCOM Mini Conference*, 2009.

[3] A. Kind, M. P. Stoecklin, and X. Dimitropoulos, "Histogram-based traffic anomaly detection," IEEE Transactions on Network and Service Management, vol. to appear, 2009.

[4] K. H. Ramah, K. Salamatian, and F. Kamoun, "Scan surveillance in internet networks," in *Networking*, 2009, pp. 614–625.

[5] V. Chandola and V. Kumar, "Summarization - compressing data into an informative representation," Knowl. Inf. Syst., vol. 12, pp. 355–378, 2007.

[6] G. Cormode and S. Muthukrishnan, "An improved data stream summary: The count-min sketch and its applications," *J. Algorithms*, vol. 55, no. 1, pp. 58–75, 2005.

[7] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and identification of network anomalies using sketch subspaces," in IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement. New York, NY, USA: ACM, 2006, pp. 147–152.

[8] M. P. Stoecklin, J.-Y. L. Boudec, and A. Kind, "A two-layered anomaly detection technique based on multi-modal flow behavior models," in *PAM: Proceedings of 9th International Conference on Passive and Active Measurement*, ser. Lecture Notes in Computer Science. Springer, 2008, pp. 212–221.

Books:

[1] Jugalkumar Kalita, Network Anomaly Detection A Machine Learning Perspectives

[2] Author- Fourzan, Data Communication

[3] Jerry v. Teplitz, Switched on Network.