ISSN No. 0976-5697

Volume 4, No. 2, Jan-Feb 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Knowledge Independent Unsuperwised Detection of Attacks on Network

Anup G. Kadu*
M.E. Imformation Technology
P.R.M.I.T.&R. Badnera,Amravati
anupkadu@gmail.com

Dr. A.S.Alvi
Information Technology
P.R.M.I.T.&R. Badnera , Amravati
Abrar_alvi@rediffmail.com

***Abstract:*** As the today's internet is ever-growing and at the same time network attacks in this huge traffic are also very rapidly increasing. It is very tedious job for network operator to detect this network attack in this growing traffic. There are many methods available for the detection of network attack but they are not efficient. Out of this many methods are working in a way of knowledge dependent fashion but this is the old way it is very time consuming. Another option to this unsupervised detection of network attack this method does not require any kind of previous knowledge about the attack and traffic. Approach comes in this paper lies under the method of unsupervised detection of network attack with the robust clustering technique.

***Keywords:*** unsupervised, knowledge independent, robust clustering.

## I. INTRODUCTION

The detection of network attack in completely unsupervised fashion is very challenging goal**.** The detection of network attacks is a paramount task for network operators in today's Internet. Denial of Service attacks (DoS), Distributed DoS (DDoS), network/host scans, and spreading worms or viruses are examples of the different attacks that daily threaten the integrity and normal operation of the network. The principal challenge in automatically detecting and analyzing network attacks is that these are a moving and ever-growing target.

Two different approaches are by far dominant in the literature and commercial security devices: signature-based detection and anomaly detection. Signature-based detection systems are highly effective to detect those attacks which they are programmed to alert on. However, they cannot defend the network against unknown attacks. Even more, building new signatures is expensive and time-consuming, as it involves manual inspection by human experts. Anomaly detection uses labelled data to build normal-operation-traffic profiles, detecting anomalies as activities that deviate from this baseline.

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 to characterize them, both in an on-line basis. The analysis is performed on packet-level traffic, captured in consecutive time slots of fixed length $\Delta T$ and aggregated in IP flows (standard *5-tuples*). 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*. [1]

## II. LITERATURE REVIEW/SURVEY

### *A. Existing Approaches for the Detection of Network Attacks:*

Paul Barford, Jeffery Kline, David Plonka and Amos Ron analyzed signals to report result of signal analysis of four classes of network traffic anomalies: outages, flash crowds, attacks and measurement failures. Data for this study consists of IP flow and SNMP measurements collected over a six month period at the border router of a large university. Their results show that wavelet filters are quite effective at exposing the details of both ambient and anomalous traffic. Specifically, they show that a pseudo-spline filter tuned at specific aggregation levels will expose distinct characteristics of each class of anomaly. [2]

Augustin Soule and co-workers developed an approach for anomaly detection for large scale networks such as that of an enterprise or an ISP. The traffic patterns they focus on for analysis are that of a network-wide view of the traffic state, called the traffic matrix. In the first step a Kalman filter is used to filter out the "normal" traffic. This is done by comparing their future predictions of the traffic matrix state to an inference of the actual traffic matrix that is made using more recent measurement data than those used for prediction. In the second step the residual filtered process is then examined for anomalies. They explain here how any anomaly detection method can be viewed as a problem in statistical hypothesis testing. [3]

Balachander Krishnamurthy proposed building compact summaries of the traffic data using the notion of sketches. He has designed a variant of the sketch data structure, *k-ary sketch*, which uses a constant, small amount of memory, and has constant per-record update and reconstruction cost. Its linearity property enables us to summarize traffic at various levels. He then implement a variety of time series forecast models (ARIMA, Holt-Winters, etc.) on top of such summaries and detect significant changes by looking for

flows with large forecast errors. He also presents heuristics for automatically configuring the model parameters.[4]

Ana L.N. Fred and Anil K. Jain explore the idea of evidence accumulation for combining the results of multiple clustering's. Initially, n d−dimensional data is decomposed into a large number of compact clusters; the K-means algorithm performs this decomposition, with several clustering's obtained by N random initializations of the K-means. Taking the co occurrences of pairs of patterns in the same cluster as votes for their association, the data partitions are mapped into a co-association matrix of patterns. This n _ n matrix represents a new similarity measure between patterns. The final clusters are obtained by applying a MST-based clustering algorithm on this matrix. Results on both synthetic and real data show the ability of the method to identify arbitrary shaped clusters in multidimensional data. [5]

## III. SYSTEM PLANNING AND DESIGN:

The proposed system will be identified to provide a solution to the problem of anomaly detection which is completely Knowledge Independent. In the Knowledge Independent Unsupervised Detection Of Network Attack. We evaluate the ability of UNADA to discover network attacks in real traffic without relying on signatures, learning, or labeled traffic. Additionally, we compare its performance against previous unsupervised detection methods using traffic from two different networks.

### *A. System Design:*

In the system design input data at first that contain the data packets. A data set is an ordered sequence of object, this may contain anomaly and we have to detect anomalies in the data set to detect that 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 .

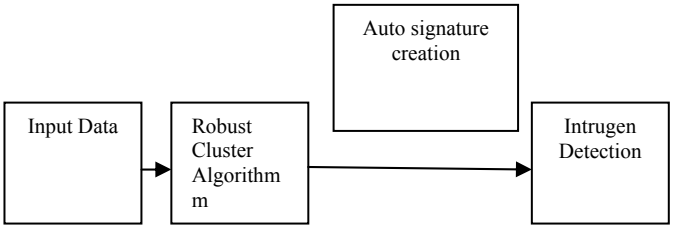


Figure.1 Organization of the system

### *B. Architecture of the Unsupervised Detection of Attacks:*

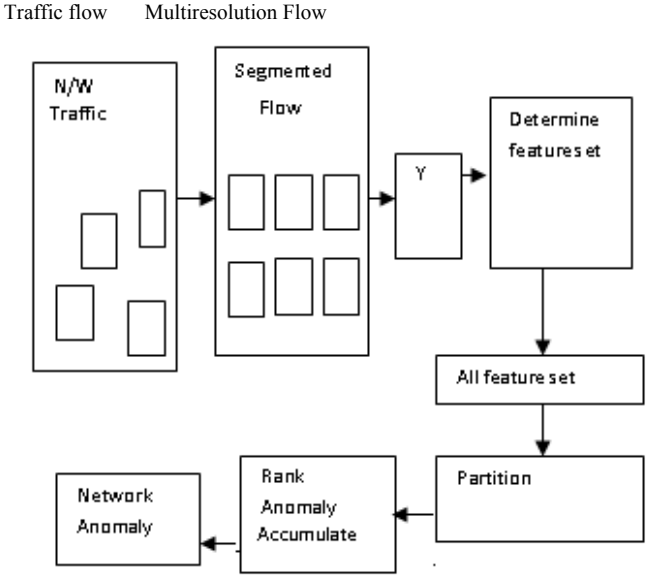


Figure 2 :Architecture of the Unsuperwised Detection Of Network Attacks

The unsupervised algorithm to detect and to automatically construct a signature for different attacks my analysis will be limited to show how the unsupervised approach can detect and characterize different network attacks without using signatures, labels, or learning.

### *C. Steps for the Unsupervised Detection of Network Attacks:*

This is an Unsupervised Network Anomaly Detection Algorithm that detects network traffic anomalies without relying on signatures, training, or labeled traffic of any kind. Based on the observation that network traffic anomalies are, by definition, sparse events that deviate markedly from the majority of the traffic, UNADA relies on robust clustering algorithms to detect outlying traffic flows.

UNADA runs in three consecutive steps, analyzing packets captured in contiguous time slots of fixed length. Figure 2 depicts a modular, high-level description of UNADA.

a. The first step consists in detecting an anomalous time slot in which the clustering analysis will be performed. To doing so first we have to break the log file in small pieces according to specific time series. And then any generic change detection algorithm is apply which is based on time series analysis to detect the anomalous time slot which then after marked as anomalous time slot.
b. In second step it takes all the input from the anomalous time slot which marked as anomalous in first step in this step, outlying flows are identified using a robust multi-clustering algorithm, based on a combination of Sub-Space Clustering (SSC), Density-Based Clustering, and Evidence Accumulation Clustering

(EAC) techniques. The evidence of traffic structure provided by this clustering algorithm is used to rank the degree of abnormality of all the identified outlying flows, building an outliers ranking.

c. In the third and final step, the outlier which is the output of outlier detection algorithm has the highest ranking are flagged as the anomalous using the simple thresh holding detection approach. As we will show through out the paper, the main contribution provided by UNADA relies on its ability to work in a completely unsupervised fashion, outperforming previous proposals for unsupervised anomaly detection.

## IV. CONCLUSION

The Unsupervised Network Anomaly Detection Algorithm presents many interesting advantages w. r. t. previous proposals in the field unsupervised anomaly detection. It uses completely unlabeled data to detect traffic anomalies, without assuming any particular model or any standard traffic flow. It detect anomaly without using previous signatures of anomalies or any kind of training by using labelled traffic. Rather using ordinary clustering techniques to identify anomalies. This approach avoids the general clustering problem. Many Unsupervised Network Anomaly Detection Algorithm have the lack of robustness of general clustering approaches, by combining the notions of Sub-Space Clustering, Density-based Clustering, and multiple Evidence Accumulation.

## V. 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.