ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

New Frontiers of Network Security: Detection of Network Attacks using Cluster

Mr.Ishraque Tanveer Khan
ME, CSE I-Year,
Department Of Computer Science and Engineering,
Everest Education Society's Group of Institutions
College of Engineering & Technology
Aurangabad, India.

Mrs.Neha Valmik
Professor,
Department Of Computer Science and Engineering,
Everest Education Society's Group of Institutions
College of Engineering & Technology
Aurangabad, India.

Abstract:

The use of network has increased over the time in critical and non-critical business transactions. This prompted hackers to attack the networks so that they get the useful information and use it for their interest, that too by illegal means. Today's attacks are succeeding far too frequently, all due to the limitations of legacy security tools. Because many security technologies developed in an earlier era of computing and networking—when attacks targeting critical information were fairly straightforward to recognize—these tools do not succeed when they cannot identify a previously unknown attack or threat vector.

Thus detection of network attacks is a critical and of utmost importance task for network operators. Huge damages are caused by network intrusion or attacks as the use of network increases for business transactions. Hence there is the need of Network Security to identify the network attacks and providing the safe and secure solutions to it.

Malicious sources are being used to attack the Networks. The classification of attacks can be done as follows: "Passive" when a network intruder intercepts data travelling through the network, and "Active" in which an intruder initiates commands to disrupt the network's normal operation. Various techniques have been suggested in the past to protect the network against all the known/unknown active & passive attacks. Upon successful authentication, the firewall applies access policies such as what services or applications the network users are allowed to access. Though quite effective to forbid unauthorized access, this component may fail to forbid potentially harmful contents such as computer worms or Trojans being transmitted over the network.

Anti-virus software or an intrusion prevention system (IPS) helps detect and inhibit the action of such malware. The methods developed so far depend on specialized signature of already known attacks or on expensive or difficult to produce labeled traffic dataset for profiling and training.

In this paper, unsupervised methods have been suggested to identify or detect the Network Attacks without following the traditional way of signatures or labeled traffic. Parallel computing highly makes the use of the clustering algorithms, which permits to perform the unsupervised detection and construction of signatures in an on-line basis.

I. INTRODUCTION:

In today's Cyber world, the very important network building-block of any Intrusion Prevention System (IPS) is the network traffic anomaly detection. The network traffic anomalies may have serious detrimental effects on the integrity, performance and efficiency of the network. Network anomalies can be anything ranging from ranging from non-malicious contents or events such as flash-crowds and failures, to network attacks such as Distributed Denial of Service (DDos), Denials-of-Service (DoS), spreading worms and network scans.

The network anomalies are moving and ever growing targets that needs to be detected as part of the network security programs. And this is the principal challenge in analyzing and detecting the traffic anomalies automatically. In the case of network attacks it is very difficult to define the set of anomalies that may arise without vagueness. The reason behind this difficulty is that new variants of new already known attacks and brand new attacking methods are continuously emerging.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

Therefore, a perfect anomalies detection system must be able to identify/detect all kinds of anomalies with diverse structure, ideally using no information at all or the minimum previous knowledge. The literature and commercial security devices are by far dominant by two different approaches: Anomalies detection and Signature based detection. The signature based detection systems are highly effective to identify the attacks which they have been instructed to alert on. And they cannot defend the network against previously unknown attacks.

The major drawback is that building new signatures is quite expensive and time consuming as it requires inspection and programming by human experts. Whereas anomalies detection uses labeled data to create acceptable operation traffic profiles and any activity that deviate from this baseline is treated as anomaly. This method can be used to detect new types of network attacks that have not been previously known. This paper presents a completely unsupervised way to detect and characterize already known as well as new network attacks, without being dependant on signatures, training, or labeled traffic of any kind.

This approach is based on robust clustering algorithms to identify known as well as completely unknown attacks, and automatically produces easy to interpret signatures on an online basis. This newly created signature can be used to detect anomalies if the attempt is made to attack the network same way.

II. UNADA (UNSUPERVISED NETWORK ANOMALY DETECTION ALGORITHM):

UNADA, an Unsupervised Network Anomaly Detection Algorithm, is widely used for knowledge-independent identification of anomalous traffic. Sub-Space-Density clustering is novel clustering technique that is employed by UNADA to identify outliers and clusters in multiple low-dimensional spaces.

A correlation-distance-based approach is used to produce an abnormality ranking of traffic flows by combining the evidence of traffic structure provided by these multiple clustering. The complete detection and characterization algorithm runs in following three successive steps:

1.The first stage consists of finding hidden attack time slot, also known as anomalous time slot.

2.In second step, the set of IP flows captured in the flagged time slot is used a input.

3.Third step is to produce filtering rules that characterize the detected attach and simplify its analysis by using the evidence of traffic structure provided by the clustering algorithms .

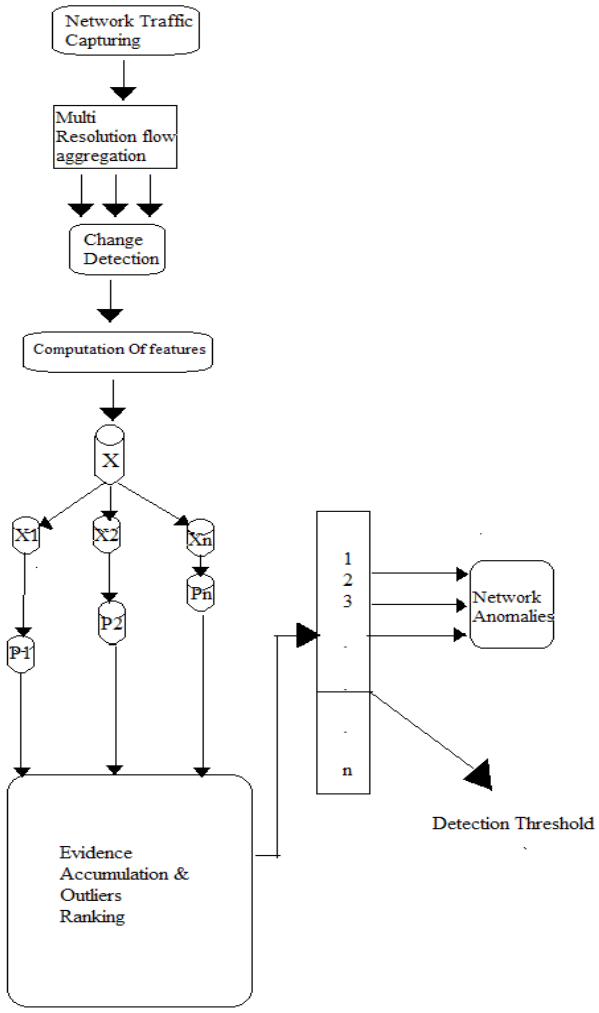

Fig. 1. High-level description of UNADA

UNADA presents several advantages with respect to current state of the art. The first and most important is that it can be directly plugged-in to any monitoring system and start to work without any knowledge or training step. It is because UNADA works in a completely unsupervised fashion.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

Second advantage is that it employs a robust density-based clustering technique that avoids most of clustering problems such as detection of particular cluster shapes, specification of number of clusters, structure masking by irrelevant features, or sensitivity to initialization. Third benefit of UNADA is that of space, it uses very-low-dimensional spaces to perform clustering. And finally, UNADA outperforms already proposed unsupervised anomaly detection methods in real world network traffic for all obvious and practical reasons.

UNADA performs in three successive stages, packets captured in contiguous time slots of fixed length are used for analysis. Figure 1 shows a high-level, modular description of UNADA. The first stage consists of indentifying an anomalous time slot where clustering analysis is to be performed.

For doing so, aggregation into multi-resolution traffic flows of captured packets is done. In the second stage, all the flows in the flagged anomalous time slot are taken as input.

And K-mean algorithm is used to identify outlying flows, in this step. The degree of abnormality of all the detected outlying flows is ranked using clustering algorithm, building an outliers ranking. The final stage consists of flagging the top-ranked outlying flows as anomalies. This step uses threshold detection approach.

UNADA uses network/port scans, spreading worms, Dos and DDoS features in the detection. Classification and its impact on the selected traffic features are defined for each unique type of attack.

All the thresholds used in the description are introduced to better explain the evidence of an attack in some of these features. DoS/DDoS attacks are characterized by many small packets sent from one or more source IPs towards a single destination IP. These attacks generally use particular packets such as TCP SYN or ICMP echo-reply, echo-request or host-unreachable packets.

Port and network scans involve small packets from one source IP to several ports in one or more destination IPs, and are usually performed with SYN packets. Spreading worms differ from network scans in that they are directed towards a small specific group of ports for which there is a known vulnerability to exploit and they generally use slightly bigger packets.

III. UNSUPERVISED ANOMALY DETECTION THROUGH CLUSTERING:

The unsupervised anomaly detection step takes as input all the IP flows in flagged time slot. At this step UNADA ranks the degree of abnormality of each of these flows, using clustering and outlier's analysis techniques. For doing so, IP flows are analyzed at two different resolutions, using either IPsrc or IPdst aggregation key. Traffic anomalies can be roughly grouped in two different classes, depending on their spatial structure and number of impacted IP flows.1-to-N anomalies and N-to-1 anomalies. 1-to-N anomalies involve many IP flows from the same source towards different destinations; examples include network scans and spreading worms/virus.

On the other hand, N-to-1 anomalies involve IP flows from different sources towards a single destination; examples include DDoS attacks and flash-crowds. IPsrc key permits to highlight 1-to-N anomalies, while N-to-1 anomalies are more easily detected with IPdst key. The choice of both keys for clustering analysis ensures that even highly distributed anomalies, which may possibly involve a large number of IP flows, can be represented as outliers. Without loss of generality, let Y = {y1, y2, .., yn} be the set of n aggregated- flows (at IPsrc or IPdst) in the flagged slot.

Each flow yi Y is described by a set of m traffic attributes or features, like number of sources, destination ports, or packet rate. Let xi Rm be the corresponding vector of traffic features describing flow yi , and X = {x1, x2, .., xn } the complete matrix of features, referred to as the feature space.UNADA is based on clustering techniques applied to X. UNADA is capable of detecting anomalies of very different characteristics. We shall therefore use k = 2 for SSC, which gives N = m (m – 1)/2 partitions.

For cluster analysis in data mining, k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes. The algorithm is applied to X. The objective of clustering is to partition a set of unlabeled elements into homogeneous groups of similar characteristics, based on some measure of similarity.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

A.Proposed 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 labelled traffic. Additionally, we compare its performance against previous unsupervised detection methods using traffic from two different networks.

B.System Design:

The first step as part of this system is to input the data that contains the data packets. A data set is an ordered sequence of objects, this may contain anomaly and we have to detect anomalies in the data set. 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 that has no previous knowledge of anomalies, the task is to detect such types of blind attacks. I am going to apply robust clustering approach for the detection of network anomalies in a completely unsupervised fashion.

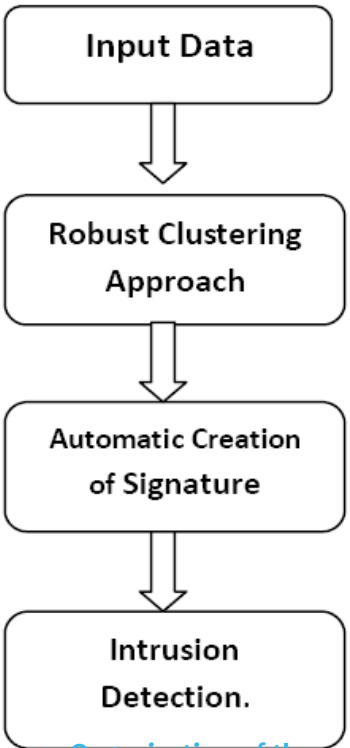

Figure 2. Organization of the System

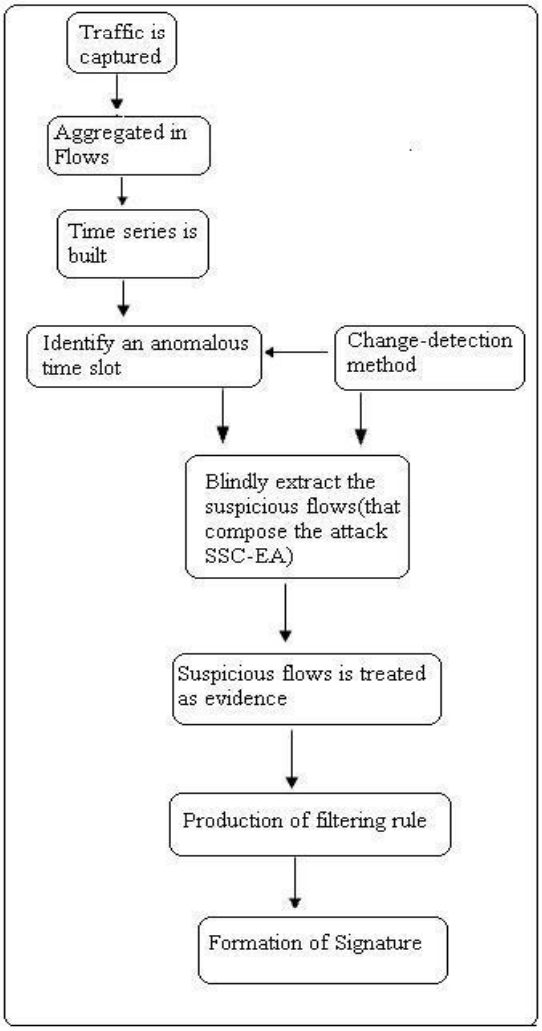

Figure 3. Workflow

This approach targets to detect both known anomalies as well as unknown attacks. This is done by the production of signature that determines the attack in an online basis algorithm that is being applied for characterizing attacks. It will run in the following stages, which is being represented by flow as shown below in figure 2. This followed the 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

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

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. 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.

C.System Implementation Steps:

Step 1:- To capture the packet of data which takes as input all the IP flows in flagged time slot by using analyzer i.e. Create Log file.

Step 2:- IP flows are additionally aggregated at different flow-resolution levels using different aggregation keys and apply sliding time windowing scheme for every 1 sec.

Step 3:- Create the feature space matrix by using following formula.x(1) = [sipadd dipadd sport dport nsipadd/ndipadd y(1)/ndipadd] Similarly, create feature space matrices (i.e. clusters) for all time windows data set i.e. X=Σ(x1,x2……xn) and then apply Clustering algorithm and declare smallest group of cluster as outlier.

Step 4:- Detect anomalies using k-means clustering algorithm, evidence accumulation and outliers ranking.

Step 5:- Create a signature. Signature will be logged and updated in the signature table. Signature table can be used in for online detection anomalous flow.

Step 6:- To detect the attack in the future this signature can ultimately be integrated to any standard security device. That is filtering rules are combined into a new traffic signature that characterizes the attack in simple terms.

IV. K-MEANS ALGORITHM:

K-means algorithm, as the underlying clustering algorithm, is used to produce clustering ensembles.

First, the data is split into a large number of compact and small clusters; different decompositions are obtained by random initializations of the K-means algorithm. The data organization present in the multiple clustering is mapped into a co-association matrix which provides a measure of similarity between patterns. The final data partition is obtained by clustering this new similarity matrix.

The main steps of K-means algorithm are as follows:-

1.Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.

2.Assign each object to the group that has the closest centroid.

3.When all objects have been assigned, recalculate the positions of the K centroids.

Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

A.An Example (K-Means):

Suppose that we have n sample feature vectors x1, x2, ..., xn all from the same class, and we know that they fall into k compact clusters, k < n. Let mi be the mean of the vectors in cluster i. If the clusters are well separated, we can use a minimum-distance classifier to separate them. That is, we can say that x is in cluster i if || x - mi || is the minimum of all the k distances. This suggests the following procedure for finding the k means:

- Make initial guesses for the means m1, m2, ..., mk
- Until there are no changes in any mean

o Use the estimated means to classify the samples into clusters

o For i from 1 to k Replace mi with the mean of all of the samples for cluster i

o end_for

- end_until

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

V. APPLICATIONS:

1.This concept can be applied on the Defence network where data security is of paramount importance. Enemies always try to crack the security by new methods to avoid suspicion. Hence this method can help detect the network without previous knowledge.

2.This concept can be used in firewall construction to detect unauthorized data.

3.In any organization such as banks, IT sector, software companies there is transmission of lot of secure data. Every bit of data is critical and important so it is necessary that it is accessible only to the authorized persons and if an unauthorized person tries to access that data then it can be detected by using unsupervised detection.

VI. CONCLUSION:

The Unsupervised Network Anomaly Detection Algorithm that we have proposed presents many interesting advantages with respect to previous proposals in the field of unsupervised anomaly detection. It uses exclusively unlabelled data to detect traffic anomalies, without assuming any particular model or any canonical data distribution, and without using signatures of anomalies or training. Despite using ordinary clustering techniques to identify anomalies, UNADA uses robust clustering technique.

An 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.

We have verified the effectiveness of UNADA to detect real single source-destination and distributed network attacks in real traffic traces from different networks, all in a completely blind fashion, without assuming any particular traffic model, clustering parameters, or even clusters structure beyond a basic definition of what an anomaly is. Additionally, we have shown detection results that outperform traditional approaches for outlier's detection, providing a stronger evidence of the accuracy of UNADA to detect network anomalies.

ACKNOWLEDGMENT:

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mrs. Neha Valmik for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project.

I would like to express my gratitude towards my mother, wife and kids for their kind co-operation and encouragement which helped me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleagues in developing the project and people who have willingly helped me out with their abilities.

REFERENCES:

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example: [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in "[3]"—do not use "Ref. [3]" or "reference [3]". Do not use reference citations as nouns of a sentence (e.g., not: "as the writer explains in [1]").

Unless there are six authors or more give all authors' names and do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Monthly Peer Reviewed Open Access International e-Journal

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

[2]P. Barford, J. Kline, D. Plonka, A. Ron, "A Signal Analysis of Network Traffic Anomalies", in Proc. ACM IMW, 2002.

[3]J. Brutlag, "Aberrant Behavior Detection in Time Series for Network Monitoring", in Proc. 14th Systems Administration Conference, 2000.

[4]B. Krishnamurthy et al., "Sketch-based Change Detection: Methods, Evaluation, and Applications", in Proc. ACM IMC, 2003.

[5]A. Soule et al., "Combining Filtering and Statistical Methods for Anomaly Detection", in Proc. ACM IMC, 2005.

[6]G. Cormode, S. Muthukrishnan, "What's New: Finding Significant Differences in Network Data Streams", in IEEE Trans. on Net., vol. 13 (6), pp. 1219-1232, 2005.

[7]G. Dewaele et al., "Extracting Hidden Anomalies using Sketch and non Gaussian Multi-resolution Statistical Detection Procedures", in Proc. SIGCOMM LSAD, 2007.

[8]A. Lakhina, M. Crovella, C. Diot, "Diagnosing Network-Wide Traffic Anomalies", in Proc. ACM SIGCOMM, 2004.

[9]A. Lakhina, M. Crovella, C. Diot, "Mining Anomalies Using Traffic Feature Distributions", in Proc. ACM SIGCOMM, 2005.

[10]G. Fernandes, P. Owezarski, "Automated Classification of Network Traffic Anomalies", in Proc. SecureComm'09, 2009.

[11]M. Ester et al., "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in Proc. ACM SIGKDD, 1996.

[12]P. Casas, J. Mazel, P. Owezarski, "Sub-Space Clustering & Evidence Accumulation for Unsupervised Network Anomaly Detection", Report LAAS-CNRS 10713, 2010.

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

[14]A. Fred, A. Jain, "Combining Multiple ClusteringsUsing Evidence Accumulation", in IEEE Trans. Pattern Anal. and Mach. Intell., vol. 27 (6), pp. 835-850, 2005.

[15]A. Jain, "Data Clustering: 50 Years Beyond K-Means", in Pattern Recognition Letters, vol. 31 (8), pp. 651-666, 2010.

[16]T. Jaakkola and D. Haussler, "Exploiting Generative Models in Discriminative Classifiers.", in Advances in Neural Inf. Processing Sys. II, pp. 487-493, 1998.

[17]L. Portnoy, E. Eskin, S. Stolfo, "Intrusion Detection with Unlabeled Data Using Clustering", in Proc. ACM DMSA Workshop, 200