Research Article Impact Factor: 0.621 ISSN: 2319-507X
Priti Dhanke, IJPRET, 2014; Volume 2 (8): 475-482 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

AUTONOMOUS NETWORK SECURITY FOR UNSUPERVISED DETECTION OF NETWORK ATTACKS

MS. PRITI B. DHANKE[1], PROF. S. P. CHHAWARE[2], MS. PRATIBHA MISHRA[3]

1. Student (M.Tech) G.H. Raisoni Institute of Engineering & Technology for women, Nagpur.
2. Assistant Professor, Priydarshani College Of Engineering, Nagpur.
3. Assistant Professor , G. H. Raisoni Institute of Engineering & Technology for women, Nagpur.

Accepted Date: 15/02/2014 ; Published Date: 01/04/2014

Abstract: The detection of network attacks is a most important 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. Signature-based detection systems are highly effective to detect those attacks which they are programmed to alert on. Anomaly detection uses labeled 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. UNADA, an Unsupervised Network Anomaly Detection Algorithm for knowledge-independent detection of anomalous traffic. UNADA uses a novel clustering technique based on Sub-Space-Density clustering to identify clusters and outliers in multiple low-dimensional spaces. The evidence of traffic structure provided by these multiple clustering is then combined to produce an abnormality ranking of traffic flows, using a correlation-distance-based approach.

Keywords: Sub-Space-Density Clustering, Unsupervised Anomaly Detection

PAPER-QR CODE

Corresponding Author: MS. PRITI B. DHANKE

Access Online On:

www.ijpret.com

How to Cite This Article:

Priti Dhanke, IJPRET, 2014; Volume 2 (8): 475-482

INTRODUCTION

The detection of network attacks is a most important 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.

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 labeled data to build normal-operation-traffic profiles. Such methods can detect new kinds of network attacks not seen before. In this paper a completely unsupervised method to detect and characterize network attacks, without relying on signatures, training, or labelled traffic of any kind. The 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 ΔT and aggregated in IP flows . IP flows are additionally aggregated at 5 different flow levels. These include source IPs, destination IPs , source Network Prefixes, destination Network Prefixes, and traffic per Time Slot.[1]

II. RELATED WORK

Most approaches analyze statistical variations of traffic volume-metrics (e.g., number of bytes, packets, or flows) and/or other traffic features using either single link measurements or network-wide data. The problem of network attacks and anomaly detection has been extensively studied in the last decade. The principal challenge in automatically detecting and analyzing network attacks is that these are a moving and ever-growing target [1]. Taxonomy allows for previous knowledge to be applied to new attacks as well as providing a structured way to view such attacks. The proposed taxonomy aims to create categories that enable this to occur easily so that similarities between attacks can be highlighted and used to combat new attacks.

A non-exhaustive list of methods includes the use of signal processing techniques (e.g., ARIMA, wavelets) on single-link traffic measurements [2], [3], and sketches applied to IP-flows [4] [7], Kalman filters [5] for network-wide anomaly detection, anomaly detection algorithm based on time-series analysis [2]–[6],PCA [8]-[9] and sketches applied to IP-flows and signature-based anomaly characterization [10].To avoid the lack of robustness of general clustering techniques, I have developed a parallel-multiclustering approach, combining the notions of Density-based Clustering [11], Sub-Space Clustering [1], and Evidence Accumulation [4]. The particular details of the algorithm are fully documented in [12]. Clustering is performed in very-low-dimensional sub-spaces, which is faster than clustering in high-dimensional spaces [3].

The Fisher Score (FS) [13], basically measures the separation between clusters, relative to the total variance within each cluster. The vast majority of the unsupervised detection schemes proposed in the literature are based on clustering and outliers detection, being [14]–[16] some relevant examples. In [14], authors use a single-linkage hierarchical clustering method to cluster data from the KDD'99 data-set, based on the standard Euclidean distance for inter-patterns similarity. In [18] reports improved results in the same data-set, using three different clustering algorithms: Fixed-Width clustering, an optimized version of k-NN, and one class SVM [16] presents a combined density-grid-based clustering algorithm to improve computational complexity, obtaining similar detection results.

III. UNSUPERVISED DETECTION OF ATTACKS

The unsupervised detection stage takes as input all the IP flows in the anomalous time slot, aggregated according to one of the different aggregation levels used in the first stage. Let **Y** = *{***y**1*, .., **y***n}* be the set of *n* flows in the flagged time slot. Each flow **y***i* is described by a set of *m* traffic attributes or *features* on which the analysis is performed. The selection of these features is a key issue to any anomaly detection algorithm, and it becomes critical in the case of unsupervised detection, because there is no additional information to select the most relevant set. To detect and characterize well-known attacks, using a set of standard traffic features. Such simple traffic descriptors permit to describe standard network attacks such as DoS, DDoS, scans, and spreading worms/virus. Let **x***i* = (*xi*(1)*, .., xi*(*m*)) R*m* be the corresponding vector of traffic features describing flow **y***i*, and **X** = *{***x**1*, .., **x***n}* the complete matrix of features, referred to as the *feature space*.[1]

IV. UNADA (Unsupervised Network Anomaly Detection Algorithm)

UNADA, an Unsupervised Network Anomaly Detection Algorithm for knowledge-independent detection of anomalous traffic. UNADA uses a novel clustering technique based on Sub-Space-

Density clustering to identify clusters and outliers in multiple low-dimensional spaces. The evidence of traffic structure provided by these multiple clustering is then combined to produce an abnormality ranking of traffic flows, using a correlation-distance-basedapproach. The complete detection and characterization algorithm runs in three successive stages.

1)The first step consists in detecting an anomalous time slot where an attack might be hidden.

2)In second stage, using as input the set of IP flows captured in the flagged time slot.

3)In the third stage, the evidence of traffic structure provided by the clustering algorithms is used to produce filtering rules that characterize the detected attack and simplify its analysis. [2]

UNADA presents several advantages w.r.t. current state of the art. First and most important, it works in a completely unsupervised fashion, which means that it can be directly plugged-in to any monitoring system and start to work from scratch, without any kind of calibration and/or training step. Secondly, it uses a robust density-based clustering technique to avoid general clustering problems such as sensitivity to initialization, specification of number of clusters, detection of particular cluster shapes, or structure-masking by irrelevant features. Thirdly, it performs clustering in very-low-dimensional spaces[9]. Finally, UNADA clearly outperforms previously proposed methods for unsupervised anomaly detection in real network traffic.[2]

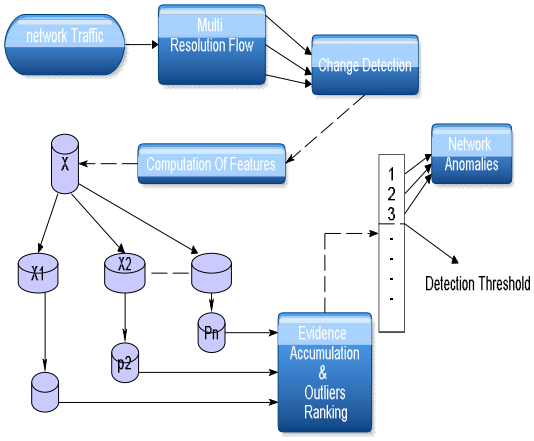

Fig. 1. High-level description of UNADA.

UNADA runs in three consecutive steps, analyzing packets captured in contiguous time slots of fixed length. Figure 1 depicts a modular, high-level description of UNADA. The first step consists in detecting an anomalous time slot in which the clustering analysis will be performed. For doing so, captured packets are first aggregated into multi-resolution traffic flows. The second

step takes as input all the flows in the time slot flagged as anomalous. This step, outlying flows are identified using a K-mean algorithm. by this clustering algorithm is used to rank the degree of abnormality of all the identified outlying flows, building an outliers ranking.In final step, the top-ranked outlying flows are flagged as anomalies, using a simple thresholding detection approach. Features used by UNADA in the detection of DoS, DDoS, network/port scans, and spreading worms. For each type of attack, we describe class and it's impact on the selected traffic features.

Table1. Features used by UNADA in the detection of attacks[2]

Type of Attack	*Impact on Traffic Features*
DoS (ICMP/SYN)	*nSrcs=nDsts=1,nPkts/sec>λ1, avgPktsSize < λ 2,nICMP/nPkts > λ 3, nSYN/nPkts > λ4.*
DDoS (ICMP/SYN)	*nDsts = 1, nSrcs > α1, nPkts/sec >α2,avgPktsSize<α3,nICMP/nPkts > α4, nSYN/nPkts > α5.*
Port scan	*nSrcs=nDsts = 1, nDstPorts >β1,avgPktsSize<β2,nSYN/nPkts > β3.*
Network scan	*nSrcs=1,nDsts>δ1, nDstPorts>δ2, avgPktsSize < δ3, nSYN/nPkts > δ4.*
Spreading worms	*nSrcs=1,nDsts>η1, nDstPorts<η2, avgPktsSize<η3, nSYN/nPkts > η4.*

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 to-wards 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.[1][2]

V. 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 outliers 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. Using 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, .., 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 , .., xn } the complete matrix of features, referred to as the feature space.UNADA is based on clustering techniques applied to X. [2]

The ability of UNADA to detect 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.

VI. CONCLUSION

The completely unsupervised algorithm for detection of network attacks. It uses exclusively unlabeled data to detect and characterize network attacks, without assuming any kind of signature, particular model, or canonical data distribution. This allows detecting new previously unseen network attacks, even without using statistical learning. 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. The use of the algorithm for on-line unsupervised detection and automatic generation of signatures is possible and easy to achieve for the volumes of traffic.

REFERENCES

1. P. Casas, J. Mazel, P. Owezarski, "Steps Towards Autonomous Network Security: Unsupervised Detection of Network Attacks", IEEE 2011 .

2. P. Casas, J. Mazel ,P. Owezarski," UNADA: Unsupervised Network Anomaly Detection using Sub-Space Outliers Ranking", UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France.

3. Rui Xu, Student Member, IEEE,D. Wunsch II, Fellow IEEE, "Survey of Clustering Algorithms", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3, MAY 2005

4. Pedro P. Casas, J. Mazel, P.Owezarski," Knowledge-Independent Traffic Monitoring: Unsupervised Detection of Network Attacks", IEEE Network January/February 2012

5. A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, B. Stiller," An Overview of IP Flow-Based Intrusion Detection", IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 12, NO. 3, THIRD QUARTER 2010.

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

7. M. Thottan and J. Chuanyi, "Anomaly Detection in IP Networks", in IEEE Trans. Sig. Proc., vol. 51 , pp. 2191-2204, 2003.

8. A. Fred, A. K. Jain, "Combining Multiple Clusterings Using Evidence Accumulation", in IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27 (6), pp. 835-850, 2005.

9. L. Portnoy, E. Eskin, and S. Stolfo, "Intrusion Detection with Unlabeled Data Using Clustering", in Proc. ACM DMSA Workshop, 2001.

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

11. G. Androulidakis, V. Chatzigiannakis, and S. Papavassiliou, "Network Anomaly Detection and Classification via Opportunistic Sampling", in IEEE Network, vol. 23 (1), 2009.

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

13. A. Fred and A. K. Jain, "Combining Multiple Clusterings Using Evidence Accumulation", in IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27 (6), pp. 835-850, 2005.

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

15. M. Ester, H. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in Proc.ACM SIGKDD, 1996.

16. J. Mirkovic and P. Reiher, "A Taxonomy of DDoS Attack and DdoS Defense Mechanisms", in ACM SIGCOMM Comput. Commun. Rev., vol.34 (2), pp. 39-53, 2004.