ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

A Survey on Abnormal User Behaviour Detection Techniques

Nandit Malviya[1], Mukta S. Takalikar[2]

P.G. Student, Department of Computer Engineering, P. I. C. T, Pune, Maharashtra, India[1]

Professor, Department of Computer Engineering, P. I. C. T, Pune, Maharashtra, India[2]

ABSTRACT: Detecting abnormal user behaviour is of great significance for a secured network, the traditional detection method, which is based on machine learning. Usually system needs to accumulate a large amount of abnormal behaviour data for training from different times or even different network environments, so the data gathered is not in line with practical data and thus affects accuracy, which increases overhead for data labelling. Network user behaviour needs to be analyzed through network flow of the network, so for accuracy of detecting and analyzing the net flow data for anomalous, user behaviour must be very efficient. The current research in this area stats algorithm like SVM, PCA are used mostly to classify data. In this paper we will be discussing the methods and process of anomaly detection based on user behaviour analytics.

KEYWORDS: Anomaly detection, Learning process, Machine learning, Security

I. INTRODUCTION

User behaviour analytics (UBA) includes tracking, collecting, assessing of user data and activities using monitoring systems. Historical data logs including network are analyzed by UBA technologies, authentication logs are collected and stored in log management, SIEM systems are used to identify patterns of traffic caused by user behaviours, both normal and malicious.[17] UBA systems are primarily intended to provide cyber security teams with actionable insights. While UBA systems don't take action based on their findings, they can be configured to automatically adjust the difficulty of authenticatingusers who show anomalous behaviour. UBA model incorporates information about: user roles andtitles from HR directories or applications, including accessing accounts and permissions; activity and geographic location of the data gathered from network infrastructure; alerts from defence in depth security solutions, and more. This data is correlated and analysed which is based on past and on-goingactivity. Normal vs abnormal behaviour profiles are developed by UBA by collecting information user behaviour activities across accounts, devices and IP addresses. Unlike signature-based threat technologies, user behaviour analytics creates a baseline for each individual user and then uses categorical, numericaland contextual information to identify anomalies and flag risky behaviour.

Abnormal user behaviour detection has always been a hot topic for network research. Thanks to the progress of machine learning, there are many different machine-learning methods been used in abnormal user behaviour detection. This technique has been used in various field to solve respective problems. Fields like medical, social networking, cyber security, cloud computing and etc. In medical [10] UBA has been applied over dataset of diabetes, cancer to find the cause of particular disease and as well as predict from the data whether its positive or negative. To decode the hacker or attacker pattern in social networking and cyber crime. Various fields comprises of different attacks and machine learning gives thebest solution of the problem. Though machine learning provides solution of the problem but it has many challenges to look upon :

- Detecting Abnormal Behaviours to Speedily Identify Insider Threats
- Real-time analysis
- Modern Threat Behaviours

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

As above given if the attacker gets into the system and start accessing credentials then its useless to detect such attack afterwards. Real-time detection of such attacks will be useful to save the system. With the evolvement of time new attacking scheme will be found by attackers so detecting those patternsare one of major challenges

II. RELATED WORK

Detection is an active security technology for the abnormal user behaviour intrusion which provides real-time protection to intercept and respond to internal attacks as well as external attacks when the network system is being jeopardized. Intrusion is defined as a collection of malicious behaviors that attempt to undermine the integrity, confidentiality or availability of resources.

You Lu*, Xutfeng Xi, Ze Hua, Hongjie Wu, NI Zhang (2014) [1] proposed a method by implementing collaborative learning with semi-supervised learning where they replaced cross validation with integration of member classifiers to reduce the overhead of labelling. By this approach they have tried to solve two challenges: Need of labelled data, overhead of labelling data.

KhurumNazirJunejo, Jonathan Goh (2016) [3] proposed nine state-of-the-art ML classifiers that are fast and scalable,though somewhat sensitive to noise.Three of classifiers represent discriminative classifiers, namely support vector machine (SVM), neural network (NN), and instance-based learning (IBK). Three other classifiers are based on decision trees, namely random forest (RF), J48, and best-first tree. The remaining three are statistical classifiers, namely naïve Bayes (NB), Bayesian network (BayesNet), and polynomial logistic regression (LR). This all work for defending system already breached and also classify which attack it was though many intrusion detection system are present which work over network layer.

K. Hanumantha Rao, G. Srinivas, AnkamDamodhar and M. Vikas Krishna (2011) [4] have described types of intrusion Detection Systems namely network intrusion detection system(NIDS), Host-based Intrusion Detection System (HIDS), Protocol-based Intrusion Detection System (PIDS). Their proposed method comprises of two algorithm working together i.e K-means and ID3 Decision trees. This method was mainly designed for two challenges-misuse of detection and anomaly detection.

Hamed Haddad Pajouh, GholamHosseinDastghaibyfard, SattarHashemi (2015) [5] have given a method where they have used naive bayes for first stage classification and for better separation between normal and anomalous activities KNN-classifier is used. They have used linear discriminant analysis (LDA) for feature reduction. Many attacks (like DoS, R2L, U2R) have been detected and generated false alarm rate
which was additional thing from past research.

N. Pandeswari, Ganesh Kumar (2015) [6] have deployed their anomaly detection process over cloud. Their cloud environment have cloudsim 3.0 installation and anomaly detection process have naive bayes and ANN-classifiers for detection. The attack types are categories such as Denial of service, Probe, R2L, U2R. They have implemented both the algorithm in term of serving as a semi-supervised approach

Kim et al. proposed anomaly detection method based on SVM [7], and evaluated its performance via KDD99 data. The main focus is on data pre-processing which mainly include data creation, feature creation, feature reduction. On the basis of pre-processing of data they have categorized anomaly detection in packet-header anomaly detection, protocol anomaly detection and content-based anomaly detection.

Laskov et al. put forward one-class SVM method for intrusion detection [8], which performed well in respect of false alarm rate. Tsang et al. held up core vector machine CVM [9], which can finish fast training based on large data set; Khan et al. combined SVM and hierarchical clustering [10]. Basically they have worked on dataset of image where they detect facial parameters or non-facial things

Robert Mitchell and Ing-Ray Chen (2015)[11] proposed model to defeat inside attackers that violate the integrity of the medical cyber physical system (MCPS) with the objective to disable the MCPS functionality and while limiting the false alarm probability to protect the welfare of patients is of utmost importance.

Jaime Devesa, Igor Santos, XabierCantero, Yoseba K. Penya, Pablo G. Bringas (2010) [12] gave a detection method where they extracted features by defining their own set of rules in form of regular expressions and these expressions includes behaviour as well family of malwares. For detection, trainingand testing ML algorithms naive Bayes, Random forest (with the forest of 100), J48 (confidence of .25), SVM (Weka model trained SVM)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

III. APPROACHES FOR ABNORMAL USER BEHAVIOUR DETECTION

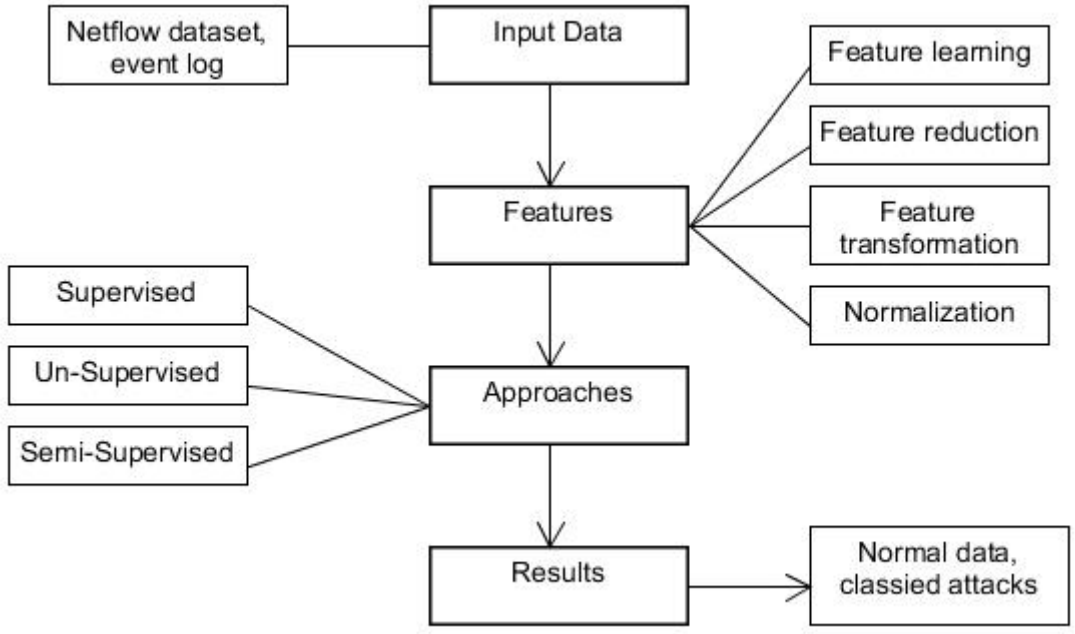

Fig 1 General Structure for Abnormal user behaviour detection

Anomaly detection is referred to the identification of items or events that do not conform to an expected pattern or to other items present in dataset. Typically these anomalous items have the potential of getting translated into some kind of problems such structural defects, errors of fraud. Using machine learning for anomaly detection helps in enhancing speed of detection. As we see in fig 1(General structure for abnormal user behaviour detection) a general flow of how machine learning algorithms works is given. For any defined algorithm most important is data in right form for that many techniques are there namely feature learning, feature reduction, feature transformation and normalization.
Intrusions are those activities that can damage information systems. Intrusion detection has been gaining broad attention. Anomaly detection can be a key for solving intrusions, as while detecting anomalies, perturbations of normal behaviour indicate a presence of intended or unintended induced attacks, defects, faults, and so on. Machine learning algorithms have the ability to learn from data and make predictions based on that data. Machine learning for anomaly detection includes techniques that provide a promising alternative for detection and classification of anomalies based on an initially large set of features.

Machine learning algorithms are broadly classified into three categories- Supervised learning, unsupervised learning and Semi-supervised learning.

A. Supervised Learning algorithm

This method requires a labelled training set that contains both normal and anomalous samples for constructing the predictive model. The most common supervised algorithms are supervised neural networks, parameterization of training model, support vector machine learning, k-nearest neighbors, Bayesian networks and decision trees.These supervised techniques have several advantages, including the capability of encoding interdependencies between

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

variables and of predicting events, along with the ability to incorporate both prior knowledge and data. In supervised learning if labels are in definite numbers and are real the problem is stated as regression problem and when labels are unordered then it is termed as classification problem.

There are various supervised models defined above and working of all models differs with the type of problem that is stated. K-nearest neighbor (k-NN) is one of the most conventional nonparametric techniques that are used in supervised learning for anomaly detection. It calculates the approximate distances between different points on the input vectors and then assigns the unlabelled point to the class of its K-nearest neighbors. The Bayesian network is another popular model that can encode probabilistic relationships among variables interest.

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.A decision tree typically starts with a single node, which branches into possible outcomes. Each of those outcomes leads to additional nodes, which branch off into other possibilities. This gives it a treelike shape.There are three different types of nodes: chance nodes, decision nodes, and end nodes. A chance node, represented by a circle, shows the probabilities of certain results. A decision node, represented by a square, shows a decision to be made, and an end node shows the final outcome of a decision path.

B. Unsupervised Learning

These techniques do not require training data. They are based on two basic assumptions. First, they presume that most of the network connections are normal traffic and only a small amount of percentage is abnormal. Second, they anticipate that malicious traffic is statistically different from normal traffic. Based on these two assumptions, data groups of similar instances that appear frequently are assumed to be normal traffic and those data groups that are infrequent are considered to be malicious. The most common unsupervised algorithms are self-organizing maps (SOM), K-means, C-means, expectation-maximization meta-algorithm (EM), adaptive resonance theory (ART), and one-class support vector machine. One popular technique is the self-organizing map (SOM). The main objective of the SOM is to reduce the dimension of data visualization.

The main idea of k-means algorithm is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated.

C. Semi-supervised Learning

Semi-supervised learning is one of the learning models getting its status in machine learning as well as artificial intelligence with its adjusting nature with unlabeled data. With such a model need of labelled data is reduced to an extent. Basically it is a combination supervised learning and unsupervised learning which can perform better than supervised learning model in some cases. As now a days unlabeled data is increasing so companies like GOOGLE, FACEBOOK are favoring semi-supervised learning so to analyze the user behavior as well activities to learn about the regularities and irregularities in the activities.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

IV. CONCLUSION

The following paper gives a brief knowledge about the techniques that can be implied in Abnormal User Behavior detection. Our paper also gives an idea of various fields where user behavior plays an important role and how analysis of user activities can help avoiding some destruction or frauds. As all models cannot be covered but the established algorithms in this field are briefly described.

REFERENCES

1. You Lu, Xuefeng Xi, Ze Hua, Hongjie Wu, Ni Zhang "An abnormal user behavior detection method based on partially labelled data" COMPUTER MODELLING NEW TECHNOLOGIES 2014 18(6) 132-141,March 2014.
2. Bi M, Xu J, Wang M, Zhou F. Anomaly detection model of user behavior based on principal component analysis. Journal of Ambient Intelligence and Humanized Computing. 7(4):547-54, 2016 Aug 1.
3. KhurumNazirJunejo, Jonathan Goh "Behaviour-Based Attack Detection and Classification in Cyber Physical Systems Using Machine Learning" CPSS 16' ACM, Xi'an, China May 30-June 03 2016.
4. Hanumantha Rao, G. Srinivas, AnkamDamodhar and M. Vikas Krishna "Implementation of Anomaly Detection Technique Using Machine Learning Algorithms", International Journal of Computer Science and Telecommunications [Volume 2, Issue 3, June 2011].
5. Pajouh HH, Dastghaibyfard G, Hashemi S. Two-tier network anomaly detection model: a machine learning approach. Journal of Intelligent Information Systems, pp: 61-74 2017 Feb 1.
6. Pandeeswari N, Kumar G. Anomaly detection system in cloud environment using fuzzy clustering based ANN. Mobile Networks and Applications.21(3):494-505 2016 Jun 1.
7. Deepaa A J, Kavitha V "A Comprehensive Survey on Approaches to Intrusion Detection System" Procedia Engineering 38 2063-9 2012.
8. Davisa J J, Clark A J "Data preprocessing for anomaly based network intrusion detection A review" Computers Security 30 353–75, 2011.
9. Tsang IW, Kwok JT, Cheung PM. Core vector machines: Fast SVM training on very large data sets. Journal of Machine Learning Research. 363-92005;6(Apr).
10. Khan L, Award M, Thuraisingham BA new intrusion detectionsystem using support vector machines and hierarchical clustering VLDBJournal 16(4) 507-21 2007
11. Mitchell, R. and Chen, R., "Behavior rule specification-basedintrusion detection for safety critical medical cyber physical systems" IEEE Transactions on Dependable and Secure Computing, 12(1), pp.16 – 302015.
12. Jaime Devesa, Igor Santos, XabierCantero, Yoseba K. Penya and PabloG. Bringas "Automatic behaviour-based Analysis and Classification Systemfor Malware Detection" Deusto Technological Foundation, Bilbao,Spain (2010)
13. Yao, H., Y. Liu, and C. Fang, "An Abnormal Network Traffic DetectionAlgorithm Based on Big Data Analysis" International Journal of Computers,Communications Control, 2016.
14. Yao, H., Y. Liu, and C. Fang, An Abnormal "Network Traffic Detection Algorithm Based on Big Data Analysis" International Journal of Computers,Communications Control, 2016.
15. Hsieh, C.-J. and T.-Y. Chan. "Detection DDoS attacks based on neural network using Apache Spark". in Applied System Innovation (ICASI), International Conference on. 2016.
16. Ambusaidi MA, He X, Nanda P, Tan Z. "Building an intrusion detection system using a filter-based feature selection algorithm". IEEE transactions on computers. 1;65(10):2986-98.2016 Oct
17. Meng Jiang and Peng Cui, Tsinghua University, Christos Faloutsos, Carnegie Mellon University, "Suspicious Behavior Detection: Current Trends and Future Directions" IEEE Computer Society January/February 2016