International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Machine Learning Based Network Anomaly Detection

Mohammad Kazim Hooshmand, Doreswamy

Abstract*—Network Anomaly Detection Systems (NADSs) play prominent role in network security. Due to dynamic change of malware in network traffic data, traditional tools and techniques are failing to protect networks from attack penetration. In this paper we propose a two-phase model to detect and categorize anomalies. First, we selected Random Forest based on the highest accuracy-score out of eleven commonly used algorithms tested with the same set of data. The RF is used to detect anomalies and generate an extra feature named "attack-or-not". Secondly we fed Neural Network with the data having "attack-or-not" feature to differentiate attack categories, which will help treating each type accordingly. The model performance was good, it scored 0.99 for both Precision and Recall in anomaly detection phase and 0.93 for Precision and 0.88 for Recall in attack categorization phase. We used UNSW-NB15 data set in our study.

Keywords: Machine Learning, Neural Network, Cyber Security, Network Anomaly Detection and UNSW-NB15.

I. INTRODUCTION

A. Problem Statement

Network security has become a necessity due to proliferation of information technology in everyday life. Anomalies are deviations from some established rule(s) in the form of patterns that do not conform with a well-known notion of normal form. Every system may have anomalous behaviors, identifying the relevance of that is called anomaly detection. Anomaly detection is one of the key task of data analysis to detect anomalous data from a given data set. It has been studied widely in machine learning and statistics, synonymously named novelty detection, outlier detection, deviation detection and exception mining [1]. Network Anomaly Detection (NAD) is a key tool in Network Intrusion Detection Systems (NIDSs), played key roles in identifying novel attacks in the last three decades and it is domain specific [1]. Despite of many research conducted in the domain, yet developing comprehensive models to cope with rapid changes of data in term of attack frequencies, types and nature of attacks (existing/zero-day-attack), are highly recommended.

Manuscript published on November 30, 2019.
* Correspondence Author
Mohammad Kazim*, Computer Science, Mangalore University, Mangalore, Karnataka, India. Email: kazimhooshmand@gmail.com
Dr. Doreswamy, Computer Science, Mangalore University, Mangalore, Karnataka, India. Email: doreswamyh@yahoo.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Detection of anomalies with high accuracy and less computational cost from large volume of data is one of the outstanding challenges for Intrusion Detection Systems (IDSs). Selecting effective features from data with the help of data mining techniques and machine learning algorithms will reduce computational cost and ultimately will boost anomaly detection with a better accuracy.

B. Cyber Security

Cyber security is set of tools, practice and guidelines to protect networked elements from attacks and unauthorized access. The aim is to avoid damages and bring safety and privacy to the hardware, software and data [2]. Collaborative efforts of cybersecurity professionals and researchers lead to designing a variety of cyber defense systems for the purpose of maintaining confidentiality, integrity and availability (CIA) of information in a cyber-environment [3]. Figure 1 shows a conventional cybersecurity system which addresses various threats.

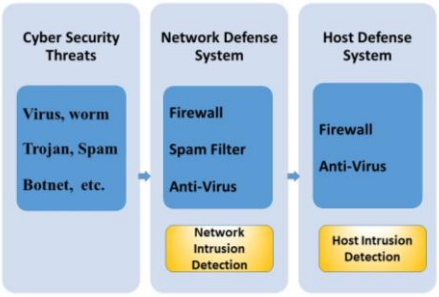

Figure 1: Conventional Cybersecurity System [3]

C. Cyber Security Solutions

1) Proactive Security Solutions: The improvement of data mining techniques and information technology brought lots of chances for media users and internet users to explore new information which can be sensitive and requires new research domain. The researchers consider data mining algorithms from the view point of privacy preservation data mining (PPDM). The PPDM algorithms can be DM techniques such as statistical methods, Bayesian Networks, unsupervised clustering algorithms etc., [3].

2) Reactive Security Solutions: In NIDSs, it is assumed that intrusion will manifest itself in a trace, the trace of normal is not the same as intrusive ones. According to detection principles, intrusion detections are classified into below categories [3].

- Misuse Detection or Signature Detection: It measures the likeliness of input events and the signature of intrusions which is already known. This method is good for detection of known attacks [3].

© Copyright: All rights reserved.

- Anomaly Detection: This method is divided into two steps, training step and detection step. In the training step, machine learning algorithms are trained in the absence of attack to generate profile for normal pattern, in detection step,
  events are labeled as attack if there is deviation from normal pattern. Figure 2 shows a generic framework of network anomaly detection (NAD). Input data needs processing, however the processing techniques vary from method to method in NADs. The broad categories of anomaly detection are supervised and unsupervised, the position of each one is shown in the below diagram. The result of anomaly detection is evaluated either by score or label [1].

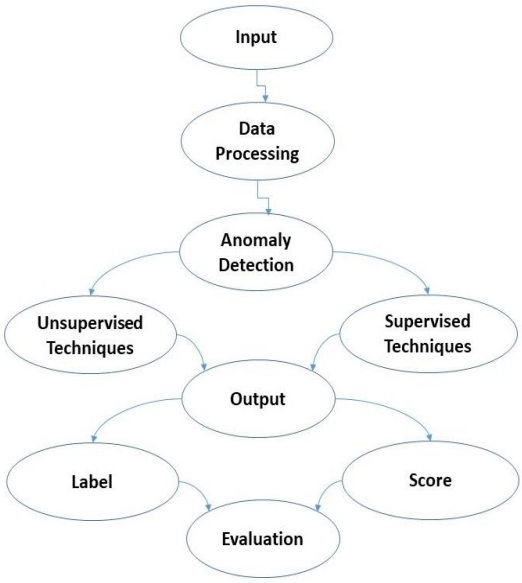

Figure 2: Generic Framework for NADs [1]

- Hybrid Detection: Both misuse detection method and anomaly detection method have drawbacks, as misuse method is not capable to handle unknown attacks and anomaly detection method usually produces false alarms. To overcome these disadvantages, a combination of the two are suggested [3].
- Scan Detection: Scan detection generates alerts if attackers scan the system, before attack is launched [3].
- Profiling Modules: It groups similar network connections and search for dominant behaviors using clustering algorithms [3].

D. Data Mining

Data mining constantly attempts to improve knowledge discovery process, it uses statistical models, AI, mathematical models and machine learning algorithms to extract previously unknown patterns from a data. The extracted pattern is helpful for distinguishing anomalous behaviors from normal behaviors in data. Data mining has two main methods i.e., supervised and unsupervised, the supervised methods use labeled data and unsupervised methods use unlabeled data as their training set. Examples of supervised methods are classification and regression and examples of unsupervised methods are association and clustering [3].

E. Machine Learning

Scientific model building process based on knowledge discovered from sample training data set is called learning. Machine learning is a complex computation process that recognizes patterns automatically and making intelligent decision based on the sample data. Methods of ML can be categorized into four i.e., connectionist-based, symbol-based, behavior-based and immune system-based. Depending on training data availability, ML is categorized mainly into supervised and unsupervised. Supervised is again divided into structure based and objective based, examples of supervised method are Artificial Neural Network (ANN), Support Vector Machine (SVM) and Decision Tree. In unsupervised method there is no target or label provided in the sample data, some of the most famous are hierarchical clustering, k-mean and self-organization map [3] [4].

The remaining parts of this paper is structured in a way that Section 2 describes UNSW-NB15 data set, Section 3 is related work review, section 4 explains methodology of the study, Section 5 is result and performance and Section 6 concludes the study.

II. DESCRIPTION OF UNSW-NB15 DATA SET

Network data set is set of events captured at connection time and the aim is evaluation of models' performance [5]. Events are divided into two categories, depending on where they come from, if it is originated with network traffic then it is network-based; if it is originated with log files then it is called host-based. Network-based events include network traffic, as example a sequence of TCP/IP, and host-based events include a sequence of commands, which are executed by users. There are different Pcap (Packet Capture) tools to capture network data and generate data sets such as Solaris BSM for SUN, LibPcap for Linux and Winpcap for Windows [3].

There are many data sets available freely for research purposes, however commonly used data set are KDD98, KDDCUP99, NSLKDD, UNSW-NB15 (University of New South Wale Network Based 2015) and many others. KDD98, KDDCUP99 and NSLKDD were generated decade(s) ago, so they do not reflect modern low footprint attacks. Based on [6] UNSW-NB15 data set was created by the IXIA PerfectStorm tool in the ACCS. ACCS is Cyber Range Lab of the Australian Centre for Cyber Security. The goal of generating this data set was to have a data set with a hybrid real modern events along with synthetic contemporary attack behaviors. Tcpdump tool was utilized to capture the raw network traffic (Pcap files). The data set generating simulation was 16 hours for first period which was conducted on 22-Jan-2015 and it was 15 hours for second simulation, which was conducted on 17-Feb-2105 for capturing 100 GBs of raw data. Due to modern attack exploitation, the IXIA tool was configured in such a way to simulate 1 attack per second during the first simulation and 10 attacks per second during the second simulation. The UNSW-NB15 data set has nine types of attacks, namely, Fuzzers, Backdoors, DoS, Analysis, Exploits, Reconnaissance, Generic, Shellcode and Worms.

© Copyright: All rights reserved.

The Argus, Bro-IDS along with twelve algorithms were used to generate additional features for the data. The data set has totally 47 features and 2 class labels. Total number of CSV records in this data set is 2,540,044, additionally there are 2 separate files training (175,341) and testing (82,332) which is more polished data. Totally 10 classes exist in the data, 9 types of attacks and normal. In this study we used CSV format of UNSW-NB15 data set which is more than 2.5 million records. The speed of flow from source to destination during the simulation has been on average 5-8.5 megabytes per second. Table I compares some commonly used data sets of network traffic.

Table I: Comparison of Popular Cyber Security Data Sets [5]

Datasets	Realistic network configuration	Realistic network traffic	Labelled observations	Total interaction capture	Full packet capture	Many malicious scenarios
KDD99	T	F	T	T	T	T[8]
NSL-KDD	T	F	T	T	T	T[8]
CAIDA	T[1]	T	F	F[5]	F[4]	F[2]
DEFCON	F	F[5]	F	T	T	T[8]
ISCX	T[2]	T[6]	T	T	T	T[10]
DARPA-2009	T[1]	T[5]	F	F[5]	T	T
UNSW-NB15	T	T[9]	T	T	T	T[10]

1. Net. config. info. unavailable.
2. Basic net. traces capture.
3. payload unavailable; most reduced trace info.
4. payload unavailable; some packet has no flag.
5. Comprises no packet contents and no host/proto. info.
6. Designed to include profiles of network info.
7. Only malicious traffic.
8. Doesn't reflect current trends.
9. Has large no. of proto. and services.
10. Has modern security events and malware scenarios.

UNSW-NB15 is labelled data, labels are cat-attack and label, cat-attack is nominal values i.e., name of attacks and label is binary value, 0 for normal and 1 for attack, so classification can be binary or multi-classification.

III. RELATED WORKS

Machine learning algorithms have been used widely by many researchers to develop network-based intrusion detection systems (NIDSs) as well as host-based intrusion detection systems (HIDSs) for classification and detection of cyber-attacks in network level and host level. As attacks are changing dynamically in term of volume and methods, so still existing studies have not shown a detailed performance analysis of various ML algorithms on various data sets [7]. We present some previous researchers' studies briefly here to understand the state of the art in the area.

Priyanka Dahiya et al [8], conducted a research to evaluate performance and effectiveness of Network Intrusion Detection Systems (NIDSs) using UNSW-NB15 dataset on Apche Spark. They used two feature reduction algorithms i.e., CCA (Canonical Correlation Analysis) and LDA (Linear Discriminant Analysis) and in addition to that, they applied seven classifiers (Rep Tree, Nave Bayes, Random Forest, Random Tree, Random Commitee, Bagging and Randomizable Filtered). The performance of 7 different classifiers were evaluated based on various parameters such as accuracy, false positive rate, training time, precision, recall and ROC area. It was found that the random tree was the winner based on various performance parameters and LDA was better feature reduction technique. They experimented classification algorithms with different data portions and the max accuracy scored through all the experiments was 95.53%. Random forest algorithm was used to reduce the risk of overfitting and also it is very efficient for large databases and consider many fewer attributes for each split. The researchers highlighted that the result with UNSW-NB15 dataset was more realistic compared to researches have been conducted for the same purpose with KDDCUP99 and NSL-KDD data sets, as those data sets were not representing modern attacks comprehensively.

Dipali Gangadhar Mogal et al [9], built a central point algorithm with respect to Association Rule Mining (ARM) as a method of feature selection to adopt the relevant features from the UNSW-NB15 and the KDD-99 data sets. They used central point algorithm to compute attribute values of central points, also applied apriori algorithm to select features using association rule mining (ARM). Logistic Regression and Naïve Bayes was used to evaluate data sets, they analyzed the result in term of detection accuracy with respect to processing time and found that CP with apriori is much better than apriori alone and also they compared the two data sets using both methods. The performance on UNSW-NB15 was much better than KDDCup99.

Mustapha Belouch et al [10], conducted an experimental evaluation to compare performance of four commonly used classification algorithms such as Naive Bayes, SVM, Decision Tree and Random Forest on Apache Spark big data environment to evaluate detection time, building time and prediction time of classifiers, the task of classification algorithms was to classify network traffic data as normal and attack. They used UNSW-NB15 data set for the evaluation and concluded that Random Forest algorithm was outperformer with respect to accuracy (97.49%), specificity (97.75%), sensitivity (93.53%) and execution time among all the other tested algorithms.

Hebatallah Mostafa Anwer et al [11], used different strategies for feature selection using filter and wrapper methods. Naive Bayes and J48 machine learning algorithms were used to classify UNSW-NB15 data set.

© Copyright: All rights reserved.

They found that the best strategy is based on 18 features applying J48 classifier and GR ranking method which got 88% accuracy and speeded up factor of 2. As their future work they suggested that ANN and SVM will be used to extend the framework under different feature selection methods along with the majority voting scheme between all classifiers to boost IDS's accuracy.

Mohamed Idhammad et al [12], studied victim-end DoS detection on artificial neural network, in this study feed forward and back propagation learning algorithms were used. Unsupervised correlation-based method was applied to select features. The experiment was conducted in three steps, first was data collection of incoming traffic, second was feature selection for DoS detection and the third was classification into normal traffic and DoS attack. Performance evaluation was tested on the two well-known data sets such as NSL-KDD and UNSW-NB15, result was reported satisfactory compared to other DoS detection methods, as feature selection reduced data dimensionality and improved training time as well as detection time.

Nour Moustafa, Jill Slay [13], developed a hybrid approach feature selection algorithm based on attribute values central points, followed by an ARM (association rule mining). First data set was divided into equal partitions to reduce processing time, output of CP was fed as input to ARM to select highly ranked features. In the decision engine Logistic Regression, Expectation-Maximization clustering and Naïve Bayes techniques were used for Network Intrusion Detection to compare and evaluate the results. The model was tested on UNSW-NB15 and NSL-KDD data sets. They found that the model was able to boost accuracy, reduce false alarm rate and shorten processing time.

Sayantan Guha et al [14], presented a cyber-system detection approach based on Multimodal Artificial Neural Network (MANN). Genetic Algorithm was used as feature selection algorithm. K-means and Convolutional neural network (CNN) clustering algorithm were employed for estimating the types of states in cyber system and to learn the features deeply. Two attack detection systems were presented i.e., attack detection for completely observable and attack detection to estimate the states type in partially observable cyber systems, NSL-KDD and UNSW-NB15 was used for the study.

Mustapha Belouch et al [15], applied two stages classifier, RepTree algorithm and protocols subset for an IDS. They used UNSW-NB15 as well as NSL-KDD datasets to evaluate performance of their approaches. Firstly, they classified incoming network traffic flow into TCP, UDP and OTHER, then classified into normal and anomaly. In the second stage a multiclass algorithm was used to classify detected anomalies into classes in order to choose appropriate action accordingly. The number of features was reduced from 40 to 20 by combination of information gain and consistency through evolutionary search method. Accuracy of detection for complete UNSW-NB15 and NSL-KDD datasets was 88.95% and 89.85% respectively.

Nour Moustafa et al [16], proposed architectural scheme for designing a threat intelligence technique for web attack, they used a methodology such that, firstly collected web data by crawling web sites, next the important features of data were extracted using Association Rule Mining (ARM) algorithm. Using these extracted features simulated web attack data, and proposing a new Outlier Gaussian Mixture (OGM) technique for detecting known as well as zero-day attacks based on anomaly detection methodology. The experiment result showed that the proposed scheme was outperformer compared to four other machine learning mechanisms in the sense of increasing detection rate and reducing FAR. The data set used for the experiment was UNSW-NB15 and web attack data.

Yiyun Zhou et al [17], suggested a Deep Feature Embedding Learning (DFEL) framework for detecting intrusions in the environment of IoT, the experiment result highlights that DFEL boosted classifiers' accuracy for cyber-attack predictions and also detection time is saved significantly.

Hossein Gharaee et al [18], proposed an anomaly based IDS model. The algorithms applied for the study were SVM and Genetic Algorithm with additional new feature selection methods. This new model invoked feature selection methods based on Genetic Algorithm with an innovation in fitness function to reduce data dimensions and at the same time boost true positive detection. The result showed that training time was reduced and accuracy was increased and false positive rate was lower.

IV. METHODOLOGY

Main steps of high level process are explained as

- Data preprocessing.
- Splitting data.
- Feature reduction.
- Training binary classifier algorithms and selecting the outperformer based on the highest accuracy, using that to detect anomalies and generate "attack-or-not" feature.
- Training neural network to predict attack categories with training set having "attack-or-not" feature and applying performance metrics to measure how well the model perform.

A. Data Preprocessing

We use full UNSW-NB15 CSV data. It has different feature types i.e., nominal, integer, binary, float and timestamp, we slice the data based on different feature types, then each type is converted to numeric accordingly, NaN values are replaced with zero if possible, removed data points otherwise because small in number. Nominal values are unified by trimming and changing to lower case and subsequently vectorized and encoded, for example a feature with four possible nominal values was converted to four features, where value 1 means that the data point belongs to that category. We did not consider timestamp features for time being in our study. After all these operations, data are merged back and normalized between (0,1).

B. Splitting The Data

The data is very unbalanced, we under-sample the data by taking much smaller sets of data per category.

© Copyright: All rights reserved.

as we have two-phase model i.e., binary classifier and multi-classifier, so the data has to be split for each one separately, either of the step must not be trained with each other's test set, otherwise it can't proof anything.

C. Feature Reduction

We have 47 features in the data set, further it increases while we apply encoding and vectorization. The features were ranked based on their importance and then top 10 were selected for the first phase of the model.

D. Training Binary Classifiers

In this step we train, test and compare accuracy of some commonly used classifiers, then select the best classifier and generate additional "attack-or-not" feature. The algorithms are Random Forest Classifier (RFC), Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), K-Neighbors Classifier (KNN), Multinomial NB (MNB), SVC (SVM), Linear SVC (LSVC), Linear Discriminant Analysis (LDA), Logistic Regression (LGR), CART and Gaussian NB (GNB). Comparison result is given in Table-III. RF scored the highest among all, so it was selected for anomaly detection. Next, we predict anomalies with RF, performance measure is given in Table-IV.

E. Training Multi-classifier

Here we train Neural Network with training set having "attack-or-not" feature in addition to previous features used in binary classification, to predict attack categories with the testing data (with generated attack-or-not feature) and apply performance metrics to measure how well the model perform in the second phase.

F. Classification Metrics Used

IDSs performance depends on conducting a confusion matrix (Table-II), it shows classification problem and the size of table depends on the number of classes included in a particular data set. Confusion matrix helps us to compare actual and predicted labels. The terms True Positive and True Negative implies correctly predicted conditions and similarly False Positive and False Negative denote misclassified ones [5].

Table II: Binary Classification Confusion Matrix

		Actual Class		
		Positive	**Negative**	
Predicted Class	**Positive**	**TP**	**FP**	$\bar{X}$
	Negative	**FN**	**TN**	$\bar{Y}$
		X	**Y**	

Accuracy: this metric is used for reflecting overall success rate of an IDS, and computed as

$$\text{Accuracy} = \frac{TP + TN}{X + Y} \quad (1)$$

Detection Rate (DR): another name is TPR or sensitivity, is the proportion of correctly classified attack instances of the total number of attack instances.

$$\text{DR (Recall)} = \frac{TP}{X} \quad (2)$$

True Negative Rate (TNR): another name is specificity, it is the proportion of correctly classified normal instances of the total number of normal instances.

$$\text{TNR} = \frac{TN}{Y} \quad (3)$$

False Positive Rate (FPR): is the proportion of normal instances of the total number of normal instances misclassified as attacks.

$$\text{FPR} = \frac{FP}{Y} \quad (4)$$

False Negative Rate (FNR): is the proportion of misclassified attack instances of the total number of attack instances.

$$\text{FNR} = \frac{FN}{X} \quad (5)$$

Other measures commonly used are: Receiver Operating Characteristics (ROC) curve, F-measure, and Precision, which are shown as below:

$$\text{F} - \text{measure} = \frac{2(Precision * Recall)}{(Precision + Recall)} \quad (6)$$

$$\text{Precision} = \frac{TP}{X} \quad (7)$$

V. RESULTS AND PERFORMANCE MEASUREMENT

We trained and tested commonly used classifiers such as RFC, DTC, CART, GBC, KNN, MNB, SVM, LSVC, LDA, LGR and GNB with the same set of data and compared their accuracy. Random Forest scored the highest (98%) among all followed by Decision Tree Classifier (97%), the result is shown in Table-III. As the RF scored the highest accuracy, so it was chosen for anomaly detection phase in the model.

Table III: Accuracy Comparison of Binary Classifiers

No.	Classifier	Accuracy	No.	Classifier	Accuracy
1	RFC	98 %	7	SVM	76 %
2	DTC	97 %	8	LSVC	76 %
3	CART	97 %	9	LDA	76 %
4	GBC	93 %	10	LGR	76 %
5	KNN	89 %	11	GNB	52 %
6	MNB	76 %			

In this step we predict anomalies and calculate Precision, Recall and F1-Score of RF. We found that the weighted average of Precision, Recall and F1-score is 0.99. The result is given in Table-IV, class 1 means that the sample is attack and class 0 means normal.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

Table IV: Anomaly Detection Results

Class	Precision	Recall	F1-Score	Support
0	1.00	0.98	0.99	2,187,456
1	0.89	1.00	0.94	290,263
Avg./Total	0.99	0.99	0.99	2,477,719

To find types of attacks, we have used neural network with sigmoid function, the result of classification is shown in Table-V

Table V: Attack Categorization Results

Class	Name	Precision	Recall	F1-Score	Support
0	analysis	0.00	0.00	0.00	268
1	backdoor	0.00	0.01	0.00	233
2	dos	0.16	0.07	0.10	11,353
3	exploits	0.07	0.41	0.11	39,525
4	fuzzers	0.15	0.49	0.23	19,246
5	generic	0.59	0.00	0.00	210,481
6	normal	1.00	0.98	0.99	2,187,456
7	reconn.	0.04	0.00	0.00	8,987
8	shellcode	0.00	0.00	0.00	152
9	worms	0.00	0.00	0.00	18
Avg.		0.93	0.88	0.88	2,477,719

VI. CONCLUSION AND FUTURE WORK

A two-phase model using Random Forest and Neural Network was developed to classify UNSW-NB15 data set into normal and attack in the first phase and to differentiate attack types in the second phase. The model scored in average 0.99 for both Precision and Recall in anomaly detection phase, result is shown in Table-IV. Class 0 means normal and class 1 means attacks. Likewise, the model scored in average 0.93 for Precision and 0.88 for Recall in attack categorization phase, result is given in Table-V, categories of attacks are represented by 0-9 numbers except 6. Number 6 is representing normal samples which is predicted almost always, however attack samples were classified mostly to 2,3,4 and 5. Overall model performance was good, especially in anomaly detection, however it needs some improvement in attack differentiation. We will focus on improvement of attack categorization in our future studies.

REFERENCES

1. M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.
2. F. Ullah and M. A. Babar, "Architectural tactics for big data cybersecurity analytic systems: A review," *arXiv preprint arXiv:1802.03178*, 2018.
3. S. Dua and X. Du, *Data mining and machine learning in cybersecurity*. Auerbach Publications, 2016.
4. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.
5. N. Moustafa, "Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic." Ph.D. dissertation, University of New South Wales, Canberra, Australia, 2017.
6. N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set," Information Security Journal: A Global Perspective, vol. 25, no. 1-3, pp. 18– 31, 2016.
7. R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. AlNemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41 525–41 550, 2019.
8. P. Dahiya and D. K. Srivastava, "Network intrusion detection in big dataset using spark," *Procedia Computer Science*, vol. 132, pp. 253–262, 2018.
9. B. B. B. Dipali Gangadhar Mogal, Sheshnarayan R. Ghungrad, "Nids using machine learning classifiers on unsw-nb15 and kddcup99 datasets," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, pp. 533– 537, 2017.
10. M. Belouch, S. El Hadaj, and M. Idhammad, "Performance evaluation of intrusion detection based on machine learning using apache spark," *Procedia Computer Science*, vol. 127, pp. 1–6, 2018.
11. H. M. Anwer, M. Farouk, and A. Abdel-Hamid, "A framework for efficient network anomaly intrusion detection with features selection,"in *Information and Communication Systems (ICICS), 2018 9th International Conference on*. IEEE, 2018, pp. 157– 162.
12. M. Idhammad, K. Afdel, and M. Belouch, "Dos detection method based on artificial neural networks," *Int J Adv Comput Sci Appl (ijacsa)*, vol. 8 no. 4, pp. 465–471, 2017.
13. N. Moustafa and J. Slay, "A hybrid feature selection for network intrusion detection systems: central points and association rules," in *Australian Information Warfare Conference*, 2015.
14. S. Guha, *Attack detection for cyber systems and probabilistic state estimation in partially observable cyber environments*. Arizona State University, 2016.
15. M. Belouch, S. El Hadaj, and M. Idhammad, "A two-stage classifier approach using reptree algorithm for network intrusion detection,"*International Journal of Advanced Computer Science and Applications (ijacsa)*, vol. 8, no. 6, pp. 389–394, 2017.
16. N. Moustafa, G. Misra, and J. Slay, "Generalized outlier Gaussian mixture technique based on automated association features for simulating and detecting web application attacks," *IEEE Transactions on Sustainable Computing*, 2018.
17. Y. Zhou, M. Han, L. Liu, J. S. He, and Y. Wang, "Deep learning approach for cyberattack detection," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 262–267.
18. H. Gharaee and H. Hosseinvand, "A new feature selection id based on genetic algorithm and svm," in *Telecommunications (IST), 2016 8th International Symposium on*. IEEE, 2016, pp. 139–144.

AUTHORS PROFILE

Mohammad Kazim Hooshmand is currently a Research Scholar, Department of Computer Science. He received Bachelor of Computer Applications (B.C.A.) from Bangalore University in 2006 and Master of Science in Information Systems (MSc. IS) from Osmania University in 2015. After completion of his post-graduation degree, he joined as lecturer in Kabul Education University in the year 2016. He joined as Research scholar in Computer Science at Mangalore University in the year 2018, still pursuing Ph.D. on Network Anomaly Detection Models Using Machine Learning Algorithms. His areas of research interest include Data mining, Artificial Intelligence, Machine learning, Deep learning, Cyber Security and Network Security.

Dr. Doreswamy is currently a Professor of Computer Science in the Department of Computer Science. He received B.Sc. and M.Sc. degree in Computer Science from University of Mysore in 1993 and 1995 respectively. After completion of his Post-Graduation Degree, he subsequently joined and served as Lecturer in

© Copyright: All rights reserved.

Computer Science at St. Joseph's College, Bangalore from 1996-1999 and at Yuvaraja's College, a constituent college of University of Mysore from 1999-2002.Then he joined as Associate Professor in Computer Science at Mangalore University in the year 2003. He was the Chairman of the Department of Post-Graduate Studies and Research in Computer Science during 2003-2005, 2008-2012 and 2016-2018 and served at various capacities in Mangalore University, as a Chairman and member of DOC, DOS and UG/PG BOS and BOE in Computer Science. Started Ph.D. programme in Computer Science and Technology in Mangalore University with effect from the academic year 2003-04 onwards. His areas of research interests include Data Mining and Knowledge Discovery, Artificial Intelligence, Data Science, Brain Computer Interface, IoT, Machine learning and Scalable Advanced Data Mining Algorithms. He has published more than 92 contributed peer-reviewed research papers at National/International Journals and Conferences. He has chaired many National and International Conferences in India. He has completed one minor research project funded by UGC. He completed two major research projects funded by UGC and DST, New Delhi, INDIA. He served as subject expert member in many committees formed by various Universities in and outside states, Karnataka State Higher Education Council. Bangalore, UGC and KPSC. Technical advisory member to many national and international conferences, editorial member to many International Journals in Computer Science and technology, Advisory members to many institutions. He received SHIKSHA RATTAN PURSKAR, Indian International Friendship Society, New Delhi, in the year 2009. He is guiding currently Eight doctoral candidates in the areas of Data Mining and Knowledge Discovery, Data Science, BCI, Big data Analysis Machine learning and scalable Advanced Data Ming Algorithms.

© Copyright: All rights reserved.