Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

Cyber Attack Detection Using Spatio Temporal Patterns

Dr S Anandamurugan[1], M Dharani [2], R G Jaiaswath[3], S Jeeva[4]

[1]Associate Professor, Department of Information Technology, Kongu Engineering College, Tamilnadu

[2,3,4] UG Students, Department of Information Technology, Kongu Engineering College, Tamilnadu

ABSTRACT:

Cyber-attacks are rapidly growing as the Internet evolves, Furthermore, the state of cyber security is risky. A brief definition of ML/DL methods is provided, as well as Deep Learning and Machine Learning approaches for intrusion detection analysis. Using temporal or thermal correlations, each of the phases was represented by papers that were indexed, read, and summarised. There are many commonly used datasets in ML and DL, and we'll discuss the challenges of using ML/DL for cyber security, as well as make some recommendations for future study. The KDD data set is divided into four categories in this paper: simple, information, host and traffic, with all data attributes being classified using the Modified Random Forest algorithm. As a result of this empirical study on the data set, the contribution of each of four categories of attributes on DR and FAR is seen, which can significantly improve the data set's suitability for achieving full DR with minimal FAR. The proposed model successfully achieved 91 percent accuracy of classification using only 12 selected features and 97 percent accuracy of classification using 36 features, while all 42 training features achieved 98 percent accuracy of classification.

Keywords: KDD- Knowledge Discovery in Database, Intrusion detection system, Modified Random Forest, Detection Rate, False Alarm Rate.

1. INTRODUCTION

1.1. CYBER-SECURITY

An interruption detection system is software that monitors a single or a group of PCs for malicious functions such as data gathering, blue inking, or weakening system conventions. The plurality of methods used in today's interruption detection systems are incapable of dealing with the dynamic and complex nature of cyber threats on computer networks. Despite the fact that effective versatile strategies like different systems of machine learning can bring about high detection rate, bring down false caution rate and sensible calculation and correspondence cost. With the utilization of information mining can bring about incessant example mining, order, grouping and smaller than normal information stream. An involved writing study of machine

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

learning and knowledge digging strategies for remote investigation in support of disruption detection is depicted in Cyber Security. The Machine learning, Data Mining techniques are portrayed, and also a few utilizations of every strategy to digital interruption detection issues.

1.2. INTRUSION DETECTION

Hackers nowadays use a variety of tools to obtain valuable information. The key goal of this intrusion detection is to provide a thorough examination of the concept of intrusion detection, its history, life cycle, different types of intrusion detection technologies, different types of threats, various instruments and strategies, testing needs, problems, and implementations.

An IDS is also known as a burglar detector. The house's lock scheme, for example, prevents it from burglary. However, if anyone removes the lock mechanism and attempts to reach the home, the burglar detector senses the lock has been opened and raises an alarm to warn the owner. Firewalls perform an excellent job by blocking incoming traffic from the Internet in order to prevent it from bypassing the firewall.

These systems track network traffic and take proactive measures to secure networks and systems. False positives and negatives are a problem with the IPS system. A false positive is an incident that triggers an IDS warning in the absence of an assault. When there is an attack, a false negative is described as an occurrence that does not trigger an alarm.

1.3. MACHINE LEARNING (M L)

Machine learning is the most promising new AI developments. They learn algorithms in a variety of applications that they use on a regular basis. Once you use a web search engine like Google or Bing to search the internet, one of the reasons it performs so well is that it has learned how to rate web pages thanks to a knows your friends' pictures, that's machine learning. Machine learning, according to Arthur Samuel, is the area of research that enables machines to learn without being directly programmed. Arthur Samuel was well-known for his checker game. Arthur was initially better than the checkers-playing software when he created it.

1.4. SUPERVISED LEARNING

This learning procedure is based on a comparison of computed and expected outputs, i.e., learning entails calculating the error and correcting the error in order to achieve the expected result. For example, if you give a supervised algorithm a data set of houses of a certain size with real prices, it would provide more of these correct responses, such as what will be the price of a new house.

1.5. UNSUPERVISED LEARNING

Unsupervised learning is described as learning on its own, based on the input pattern, by finding and adopting. This learning is known as a clustering algorithm since the data is separated into various clusters. Google News is an example of a site that uses clustering (URL

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

news.googIe.com). Google News collects new content from around the internet and organizes them into news stories.

1.6. REINFORCEMENT LEARNING

The quality of reinforcement learning is how an agent can behave in a given situation in order to optimize some notion of long-term reward. Right output is rewarded, whereas incorrect output is punished. In contrast to the supervised learning dilemma, right input/output pairs are never presented, and suboptimal behaviors are never directly corrected in reinforcement learning.

2. LITERATURE REVIEW

2.1. IN THE DIRECTION OF CREATING A NEW INTRUSION DETECTION INTRUSION TRAFFIC CHARACTERIZATION AND DATASET

Iman Sharafaldin, has suggested in his paper that as computer networks and established applications expand in size exponentially, the substantial increase in the possible harm that can be caused by launching attacks becomes apparent. Anomaly-based approaches in intrusion detection systems fail to deploy, analyze, and test accurately due to a lack of sufficient dataset.

2.2. A DATASET FOR INTRUSION DETECTION AND AN EVALUATION FRAMEWORK

Amirhossein Gharib et al., The rising number of security risks on the Internet and data networks necessitates highly secure security solutions, according to the author of this article. Meantime, intrusion detection systems and intrusion prevention systems give a crucial role in the design and creation of a secure network infrastructure capable of detecting and blocking a wide range of attacks.

2.3. TIME-RELATED FEATURES IN THE CHARACTERIZATION OF ENCRYPTED AND VPN TRAFFIC

Gerard Draper Gil et al., It's a challenging challenge because of the constant development and generation of new applications and utilities, as well as the expansion of secure communications. Virtual Private Networks are a form of encrypt networking service that is common as a way to circumvent censorship and gain access to geographically restricted services.

2.4. FOR DETECTING THE NETWORK INTRUSION DETECTION USING UNSW-NB15 (A COMPREHENSIVE DATASET)

Moustafa et al., according to the authors, it is the lack of a comprehensive network dataset that can represent current traffic scenarios, a wide variety of low footprint trespasses, and scope

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

structured knowledge about network traffic. KDD98, KDDCUP99, and NSLKDD data sets were developed a decade ago to test network intrusion detection systems research efforts.

3. ANALYSIS OF SYSTEM

3.1 EXISTING SYSTEM

Existing cyber defence capabilities used in delivery networks are still vulnerable to cyber-attacks. To allow future energy delivery systems to reliably identify, dynamically respond, effectively withstand, and reject a cyber-attack, the development of cyber-resilient DMS functions and cyber security technologies is critical. Orthodox cyber-attack identification approaches, such as naive Bayes classifiers (BCs), depend heavily on the normality assumption**.**

3.1.1. DISADVANTAGES OF EXISTING SYSTEM

Synonyms and homonyms can create ambiguity in a term frequency approach. Depending on the target language, it can even necessitate complex pre-processing.

It cannot be used where the contents of the communications are largely non-textual data. The "words" created by mentions, on the other hand, are distinct, do not need any preprocessing (the information is frequently isolated from the contents), and can be accessed regardless of whether the contents exist.

Threat Detection through Graph Learning and Psychological Context is very tedious and accuracy rate is low.

3.2. PROPOSED SYSTEM

We want to use a flexible BC to incorporate the spatiotemporal trends of device measurements for cyber-attack detection. Spatiotemporal patterns are captured by the generalised graph Laplacian (GGL) matrix for device measurements using the RF (Random Forest Algorithm). They are used as input variables for the proposed modular BC's training method, while the marks of cyber-attack models are used as output variables.

The suggested flexible BC is used to bring the online spatiotemporal patterns captured by GGL into the testing phase, which then produces the cyber-attack detection data. To classify spatiotemporal patterns of device measurements, an unsupervised machine learning approach called GGL (Generalized Graph Laplacian) is used.

Negative selection-based detection generation has been used as inspiration for the proposed technique. The NSL-KDD dataset, which is a tweaked version of the commonly used KDD CUP 99 dataset, is used to validate this technique. The analysed parameter value is automatically chosen according to the used testing dataset to improve its adaptability and versatility. By improving clustering, you can also reduce the time it takes to generate a detection.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

3.2.1. ADVANTAGES

The recommended solution is not based on the textual nature of social media posts, is immune to rephrasing, and can be used in circumstances where the material contains non-textual media such as images, video, and audio.

The suggested link-anomaly-based methods worked better than the keyword-based methods on the KDD-CUP data sets. Find anomalous users fast and conveniently. Accuracy is high. Calculation time is kept to a minimum.

4. MODULE DESCRIPTION

4.1. SPATIOTEMPORAL PATTERNS MODULES

Chart learning approaches will quantitatively speak to spatiotemporal examples as an unaided AI methodology. The GGL, for example, will retain all edges with positive loads and, for all intents and purposes, present a wider network due to negative loads. The Lagrangian progression problem can be used to evaluate the GGL grid.

4.2. CYBER ATTACK DETECTION MODULE

Ordinary innocent BCs are generally dealt with by discretization and agree that they obey a Gaussian dissemination by using spatiotemporal examples as data sources. Regardless, this apprehension based on mathematical properties can't be applied to any of the spaces.

The developed adaptable BC, which is based on nonparametric component evaluation and does not require any ordinary suspicion, beats in a variety of spaces.

Similarly, the adaptable BC will keep track of each consistent property estimate it encounters throughout the preparation process.

4.3. EVALUATION METRIC MODULE

TPR is characterised as the percentage of detected cyberattacks (TP) that are currently identified in real-world device measurements compared to the total number of cyberattacks (NA).

A suite of metrics for the performance assessment of cyberattack detection can be obtained from the contingency table, including the probability of detection (POD), critical success index (CSI), frequency bias score (FBIAS), and success ratio (SR).

5. PROCEDURE AND ENVIRONMENTAL SETUP

Step 1): To represent spatiotemporal examples of structure estimations, a single AI technique, specifically GGL, is used.

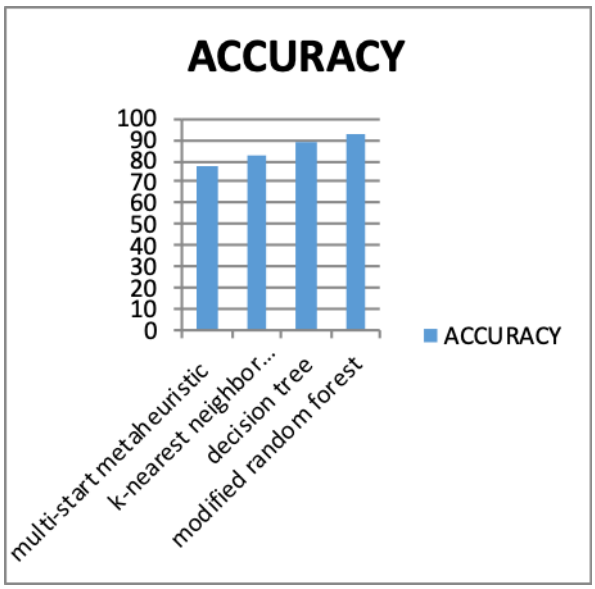
Step 2): To plan the spatiotemporal examples portrayed by the GGL network, an administered AI technique (i.e., adaptable BC) is employed.

Step 3): To test the presentation of different dentification methods, two sets of scales, namely the genuine positive rate (TPR) and the probability table, are used.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

6. RESULT ANALYSIS

Different execution metrics, are used to evaluate the procedures during their execution. The K-nearest neighbour, Decision tree, multi-start metaheuristic, and Modified Random Forest are some of the calculations covered in this paper. The accuracy rate of the multi-start metaheuristic is about 77%. KNN outperforms multi-start metaheuris with an accuracy of 82%. The accuracy of the Decision Tree is about 89 percent, which is higher than both the multi-start metaheuristic and the KNN. The accuracy of Modified Random Forest is about 93 percent, which is higher than the accuracy of all other algorithms used, such as multi-start metaheuristic, KNN, and Decision Tree.



According to the precision, the modified random forest algorithm is more accurate in detecting attacks that occur in systems

ALGORITHM	ACCURACY
Multi-start metaheuristic	77
k-nearest neighbour (KNN)	82
Decision tree	89
Modified 0random forest	93

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1439 - 1447
Received 15 April 2021; Accepted 05 May 2021.

7. CONCLUSION

The suggested mention model was combined with the MRF change-point recognition algorithm. Signature-based detection is more reliable than anomaly detection and has a lower false positive rate, but it can only detect proven attacks. Anomaly detection is capable of detecting unknown attacks, but it has a high rate of false positives.

When it comes to detecting network threats, the Intrusion Detection System is crucial. Two of the approaches used in IDS are signature-based systems and anomaly-based systems. On the other hand, signature-based systems can only detect known attacks and cannot detect unknown attacks, while anomaly-based systems can detect unknown attacks. An anomaly-based scheme is defined using an automated approach using the multi-start metaheuristic framework.

The multi-start metaheuristic has a 77 percent accuracy, KNN has an 82 percent accuracy, Decision Tree has an 89 percent accuracy, and Modified Random Forest has a 93 percent accuracy, so we conclude that Modified Random Forest is successful for detecting attacks

REFERENCES

[1] Ghorbani, Lashkari,A.H ,Sharafaldin, I, and A.A, “Toward Generating a New Intrusion Detection Dataset andIntrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, (2018).

[2] I., Lashkari, A.H. , Ghorbani, Gharib, Sharafaldin, A.A., “An Evaluation Framework for Intrusion Detection Dataset”. 2016 IEEE International Conference Information Science and Security (ICISS), pp. 1-6, (2016)

[3] A.H., Mamun M. and Ghorbani, A.A., Gil, G.D., Lashkari “Characterization of encrypted and VPN traffic using time-related features. In Proceedings of the 2nd International Conference on Information Systems Security and Privacy, pp. 407-414, (2016).

[4] Moustafa, J., N. Slay, “The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 dataset”. Information Security Journal: A Global Perspective, 25(1-3), pp.18-31, (2016).

[5] Slay, Moustafa, J, “UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). IEEE Military Communications and Information Systems Conference (MilCIS), pp. 1-6, (2015).

[6] Gurunath Chavan, Pongle, Pavan "A survey: Attacks on RPL and 6LoWPAN in IoT." IEEE International Conference on Pervasive Computing, (2015).

[7] Won Woo R,Doohwan, Deokho Kim, "A malicious pattern detection engine for embedded security systems in the Internet of Things." Sensors, pp, 24188-24211, (2014).

[8] Patil, Mangrulkar, “Network Attacks and Their Detection Mechanisms: A Review”. International Journal of Computer Applications, 90(9), (2014).

[9] Pastrone, C., Spirito, M. A., & Vinkovits, Kasinathan, P., M. “Denialof-Service detection in 6LoWPAN bas Communications Surveys and Tutorials, 12(3), pp.343-356, (2010)

[10] C.S. and Lee, S, Amin, S.O., Siddiqui, M.S., Hong, “RIDES: Robust intrusion detection system for IP-based ubiquitous sensor networks”. Sensors, 9(5), pp.3447-3468, (2009).

[11] Farooqi, Ashfaq Hussain,E.J., Kim, J.H."Attack model and detection scheme for Botnet on 6LoWPAN". In Asia-Pacific Network Operations and Management Symposium, pp. 515-518, (2009).
[12] Farrukh Aslam Khan, Farooqi, Ashfaq Hussain, "Intrusion detection systems for wireless sensor networks: A survey." Communication and networking, pp. 234-241, (2009).
[13] ed Internet of Things." In IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications, pp. 600-607, (2013).
[14] Fukuda, K. and Sugawara, T, Kanda, Y., Fontugne, R."Anomaly detection method using entropy-based PCA with three-step sketches". Computer Communications, 36(5), pp.575-588, (2013).
[15] H.Muthukrishnan,S.Akila "Performance Analysis of Implicit Trust Based Security in AODV Routing Protocol" i-manager's Journal on Wireless Communication Networks (International), Vol. 4 Issue No. 1, April – June 2015
[16] H.Muthukrishnan,Dr.S.Anandamurugan "Light Weight Security Attack in mobile Ad Hoc Network (MANET)", International Journal of Computer Sciences and Engineering(IJCSIT)(International), Vol.2, Issue-8, E-ISSN: 2347-2693, pp.56-61- 2014
[17] H.Muthukrishnan, S.Akila "Performance Analysis of Implicit Trust Based Security in OLSR Routing Protocol", i-manager's Journal on Wireless CommunicationNetworks(International), Vol Dec 2014 pp.18-24 2015
[18] H.Muthukrishnan,S.ShanthiPriya "Energy aware span routing in Adhoc Networks", CIIT International Journal of wireless Communications (International), Vol. 8, No.1, pp. 11-16 – 2016
[19] D.Z.Shen, Li, L., Yang, F.C., "A novel rule-based Intrusion Detection System using data mining". 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Vol. 6, pp. 169-172, (2010).
[20] R., Morariu, C., Pras, A., Stiller, B, Sperotto, A, "An Overview of IP Flow-based Intrusion Detection". IEEE
[21] H.Muthukrishnan, B.Sunita, S.Najeera banu, V.Yasuvanth " Observational study of WPAN and LPWA Technologies for various IoT devices and its applications " International Journal of Advanced Science and Technology, Vol. 29, No. 5, p.no 4231-4243, May 2020
[22] Hina S, Shaikh A, Sattar SA. Analyzing diabetes datasets using data mining. J Basic Appl Sci. 2017;13:466–71
[23] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J. 2017;15:104–16.
[24] H. Muthukrishnan, C. P. Thamil Selvi, Dr. M. Deivakani, V. Subashini, Savitha N. J., S. Gowdham Kumar, "Aspect Based Sentiment Analysis for Tourist Reviews" International Journal on Annals of the Romanian Society for Cell Biology, Vol. 25, Issue 3, p.no 5183 – 5194, March 2021
[25] Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet. 2009;373:1773–9.

[26] Oliver F, Rajendra AU, Ng EY, KwanHoong N, Jasjit SS. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. J Med Syst. 2012;36(1):145– 57.https://doi.org/10.1007/s1091 6-010-9454-

[27] H.Muthukrishnan "Advent of Disruptive Technologies – Assimilation of Blockchain and IoT and it's Challenges n relevance for the upliftment of Digital Relationship " International Journal of Scientific and Technology Research ISSN : 2277-8616, Volume 9, Issue 4, p.no – 672-676, April 2020

[28] American Diabetes Association. Standards of medical care in diabetes—2011. Diabetes Care. 2011;34(Suppl 1):S11–61.https ://doi.org/10.2337/dc11-S011.

[29] Anna V, van der Ploeg HP, Cheung NW, Huxley RR, Bauman AE. Socio-demographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. Diabetes Care. 2008;31(12):2288–93.

[30] Patil S, Kumaraswamy Y. Intelligent and effective heart attack prediction system using data mining and artificial neural networks. Eur J Sci Res. 2009;31(2009):642–56.