International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Cybersecurity with ML: A Multi-Algorithm Approach to Anomaly-Based Intrusion Detection

Indira P. Joshi[1], Dr. Vijaya K. Shandilya[2]

Submitted: 07/02/2024 **Revised**: 15/03/2024 **Accepted**: 21/03/2024

Abstract: In the era of escalating cyber threats, the significance of robust intrusion detection systems (IDS) cannot be overstated. Traditional methods often struggle to keep pace with the evolving tactics of malicious actors. This paper presents a novel approach to enhancing cybersecurity through the integration of machine learning (ML) techniques within anomaly-based intrusion detection systems. Specifically, we propose a multi-algorithm framework that leverages the complementary strengths of various ML models to effectively identify diverse cyber threats. Our approach aims to address the limitations of single-algorithm systems by combining the capabilities of multiple classifiers. We demonstrate the efficacy of our methodology through extensive experimentation on real-world network traffic scenarios. Results indicate that our multi-algorithm approach outperforms traditional single-algorithm solutions in terms of detection accuracy, false positive rates, and scalability. Furthermore, we discuss the practical implications of our framework in bolstering cybersecurity defenses across diverse organizational contexts. Overall, this research contributes to the advancement of anomaly-based intrusion detection systems by offering a robust and adaptable ML-driven solution capable of effectively combating emerging cyber threats.

Keywords: *IDS (Intrusion Detection System), Anomaly Detection, False Alarm Rate (FAR) and Machine Learning algorithms.*

1. Introduction

For decades, studies have been focused on PC community safety. The firm has found out the need of data & community safety technology in safeguarding its statistics. A safety attack or incursion is any deliberate or unintended try and undermine the provision, integrity, or confidentiality of any information useful resource or the facts itself. Industries are constantly being subjected to new attacks. IDS is one technique for this issue's resolution.

One approach employed by the IDS to identify assaults is gadget getting to know. The advent and advancement of algorithms and strategies that allow PC systems to independently accumulate and integrate records for you to always enhance them to do obligations successfully and efficaciously is referred to as gadget learning. System getting to know intrusion detection structures have ended up increasingly more accurate and effective in spotting new assaults in recent years. IDS is a relaxed manner that appears for numerous sorts of attacks. They are the group of methods used to locate suspicious conduct on both hosts and networks.

2. Taxonomy Of Anomaly Detection

In the past, a number of taxonomies for intrusion detection techniques have been been up, but none of them are still widely recognized. The taxonomy provided here uses six criteria to categorize IDSs, as shown in Fig. 1, and is predicated on the synthesizing of several other taxonomies. At the moment, the two most common analytical approaches of detection are cryptography and anomaly-based [1],[2]. The signature-based approach, often referred to as abuse detection, searches for a certain signature to match, which would indicate an intrusion. The shortcoming of signature-based systems for intrusion detection is their inability to recognize novel attack types or novel modifications of established attack patterns, even though they can detect most or all known assault patterns.

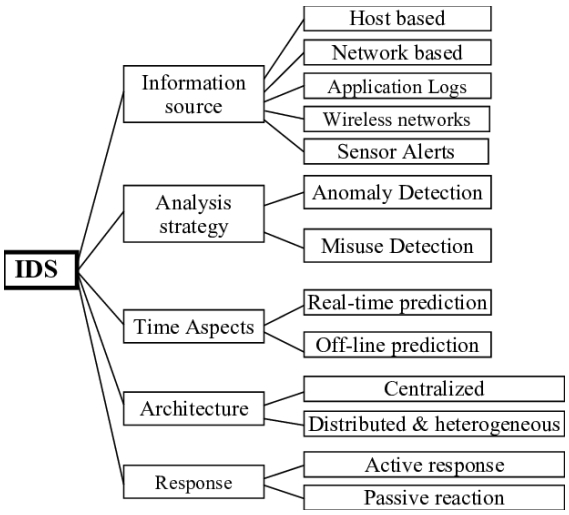

Fig 1. IDS Taxonomy

[1,2]*Department of Computer Science and Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India.*
1ipj.indira@gmail.com
2vkshandily14@gmail.com

Fig 2. Anomaly Detection classification

Anomaly detection is but any other helpful approach for intrusion detection. Since it has become first cautioned in [3], anomaly detection related to vulnerability scanning and facts protection has been the point of interest of research. In anomaly-based IDSs, the typical system in addition to network traffic behavior is depicted, and any conduct that deviates above a positive threshold is recognized as an uncommon hobby. Then again, as compared to IDSs based on signatures, anomaly-based IDSs create a larger amount of fake positives. How those technologies ought to gain knowledge of, or the way to decide what constitutes normal conduct of a platform or network environment, and a way to computerize represent this conduct, is an important topic in anomaly-based IDSs.

Anomaly detection methods may be divided into 3 primary corporations [4]: statistical, know-how-primarily based, and gadget studying-based totally, depending on the type of processing related to the target device's "behavioral" version. The intrusion prevention methodologies and comparisons of numerous processes are mentioned in [18] alongside the benefits and drawbacks of each.

2.1. Statistical anomaly-based totally IDS

A statistically anomaly-primarily based IDS monitors regular community activity, consisting of the bandwidth, commonly utilized, the protocols used, and the ports and devices which can be usually connected to each other. It notifies the administrator or person whilst extraordinary site visitors is recognized (now not everyday) [5] [7]. It's far all over again divided into time series, multivariate, and univariate models. The permissible range of values for every variable is described by using modeling the univariate version parameters as impartial Gaussian random variables. The affiliation between one or greater variables is taken into consideration by means of the multivariate model. The time collection version includes the order and inter-arrival instances of the observations, as well as their values which might be labeled as anomalous, and employs an c programming language timer, along side an invitational counter or aid degree.

2.2. Knowledge-Primarily Based Techniques

Knowledge-based structures preserve the music of difficulty-unique statistics. A format that enables the inference engine to execute deduction on the statistics in information-based totally statistics contains symbolic representations of professional's standards of judgment. One of the know-how-based totally IDS strategies this is most regularly hired is the expert system method. The 3 classes of information-based totally methodologies include expert systems, rule-based totally models, and body-primarily based models. A modified model of grammar-based manufacturing regulations are rule-based policies. A full corpus of predicted data and behaviors is localized right into a single structure through a body-based version. 3 levels are worried inside the expert machine classification of the audit records according with a set of policies. The education information is first used to become aware of diverse properties and classifications.

2.3. Machine learning-primarily based IDS

The muse of machine gaining knowledge of procedures is the introduction of an both specific or implicit model. The need for categorizing statistics and teaching the modeling technique, a procedure that lays heavy needs on resources, is a completely unique feature of these approaches. The application of statistical methods and gadget studying standards often overlaps, in spite of the latter's emphasis on growing a version that enhances overall performance based on prior statistics. As an end result, system studying for IDS has the functionality to adjust its execution method. This function could make it attractive to utilize such systems in all situations.

3. Intrusion Detection and Machine Learning

Using machine learning strategies for intrusion detection involves developing a model routinely from training statistics. Every of the information times in this set may be represented by a hard and fast of homes (functions) and the labels that go along with them. The qualities may be continuous or categorical, for instance. The suitability of anomaly detection approaches depends on the characteristics' nature. As an example, distance-based processes are commonly unsatisfactory when carried out to specific variables seeing that they have been initially designed to characterize with non-stop features. Normally, labels for facts times take the form of binary

values, along with normal and peculiar. As opposed to the use of the time period "anomaly," different researchers have used other sorts of attacks such DoS, U2R, R2L, and Probe. Learning strategies can provide greater details about the extraordinary varieties of anomalies in this way. However, experimental findings demonstrate that existing mastering methodologies is insufficiently accurate to pick out the particular types of abnormalities. Obtaining a efficaciously categorized statistics collection this is representative of all sorts of sports is relatively steeply-priced because labeling is often accomplished manually by way of human specialists. For this reason, 3 working modes for anomaly detection techniques are described: supervised learning, unsupervised gaining knowledge of, and semi-supervised learning.

4. System Design For Intrusion Detection System

1. Preprocessing Phase

With the usage of packet sniffing equipment (which includes Wireshark and Capsa), packet characteristics including IP/TCP/ICMP headers are extracted from every packet in the course of this segment of packet shooting and packet analysis. The packet header will then be divided with supply addresses, vacation spot addresses, and many others. There are sure techniques needed in this step for the selection of key features. figuring out the packet's normality or incursion, and so forth. on this stage, packets are captured from datasets (such the KDD dataset/NLS KDD) with the intention to act because the IDS's records source.

1.1. Classification

Utilize the data from the preceding step throughout the classification phase to decide if the packet is a regular packet or an attack packet. The corresponding algorithms will categorize the packet into related organizations based on the character values. Solution training and packet functions are offered at some point of the schooling section so that it will help in the improvement of rules governing mapping domain names. These tips can be modified or amended based on additional schooling. Each set of rules has a completely unique categorization approach. Untrained records are dispatched to the system all through the testing segment with a purpose to pattern whether or not proper responses are returned. Without defining the reaction class, the device operation is executed whilst accepting input packets.

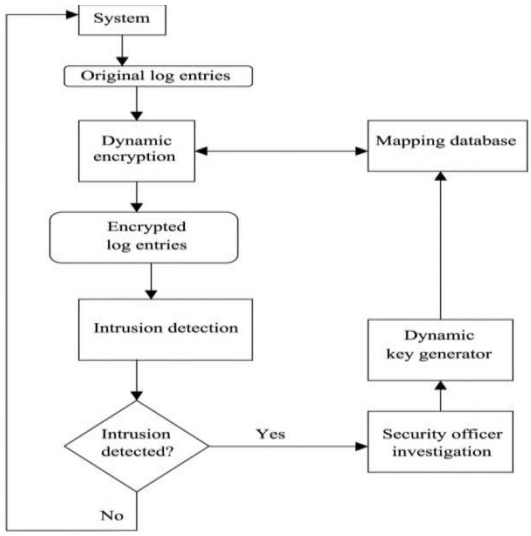

Figure 3. System Design for IDS

2. ***Post Processing***

The preprocessing output is in comparison to the answer elegance, and device performance is calculated as the sum of accuracy and false alarms. true superb, proper negative, false superb, and false bad, respectively.

1.2. Reducing False Alarms

Greater training is needed if the machine is still emitting some fake indicators throughout all the algorithms. The machine will retain to examine on its personal, loose human intervention, in line with the machine learning approach. Consequently, there may be no need for upgrading. Different classification machine learning algorithms are tested on real time data set And as shown in below fig. it shows decision tree achieve 98% accuracy among the other classifiers as LR, RF, NB, Adaboost and MLP.

V.Performance Evaluation Metric And Results

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Parameters	Definition
True Positive (*TP*) or Detection Rate (*DR*)	Attack occur and alarm raised
False Positive (*FP*)	No attack but alarm raised
True Negative (*TN*)	No attack and no alarm
False Negative (*FN*)	Attack occur but no alarm

Different classification machine learning algorithms are tested on real time data set And as shown in below figures.

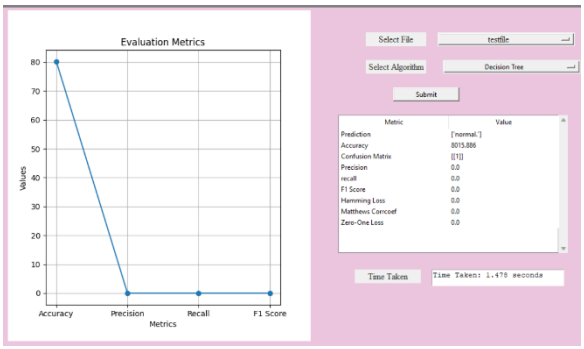

Fig.4 Evaluation of Decision Tree

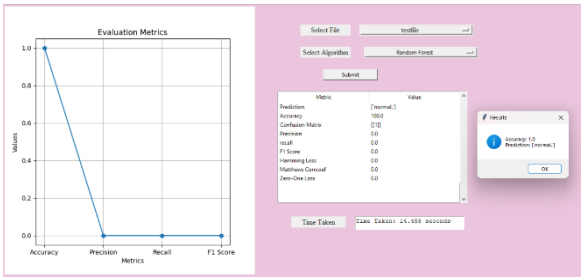

Fig.5 Evaluation of RF

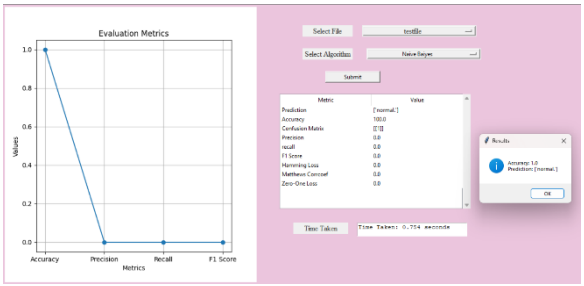

Fig.6. Evaluation of Naïve Bays

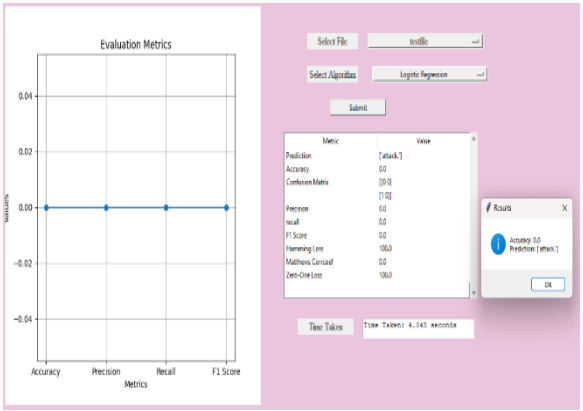

Fig.7 Evaluation of Logistic Regression

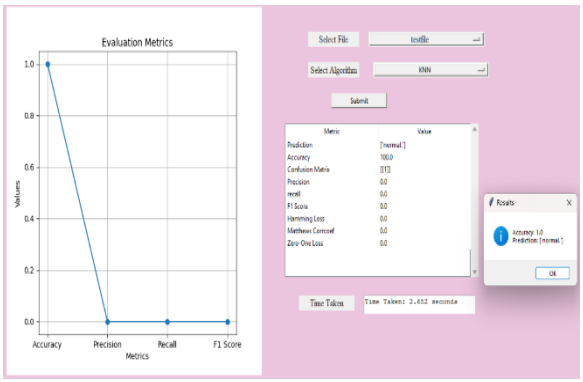

Fig.8 Evaluation of KNN

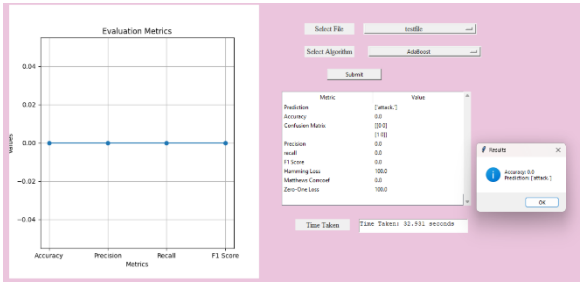

Fig.9 Evaluation of Adaboost

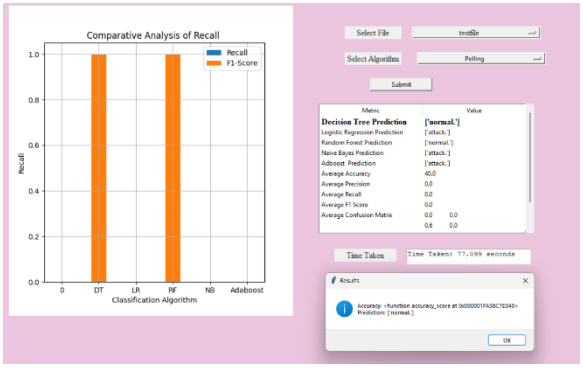

Fig.10. shows multi-algorithm Approach result.

5. Conclusion

In this exploration of "Enhancing Cybersecurity with ML",we have delved into the intersection of machine learning and cybersecurity, focusing specifically on anomaly-based intrusion detection. Through the integration of multiple ML algorithms, our study demonstrates the potential to significantly enhance cybersecurity measures by detecting intrusions effectively. By leveraging a multi-algorithm approach, we have showcased the versatility and adaptability of machine learning techniques in indentifying network anomaly. This comprehensive strategy not only enhances the detection capabilities of Intrusion Detection Systems but also strengthens the overall resilience of cybersecurity framworks against evolving thrats.

References:

[1] MananJ, Ahmed A, Ullah I, Merghem-Boulahia L, Gaiti D (2019) Distributed intrusion detection

scheme for next generation networks. J Netw Comput Appl 147.

[2] Nadiammai G, Hemalatha M (2014) Effective approach toward Intrusion detection system using data mining techniques. Egypt Inform J 15:37–50.

[3] Almseidin M, Alzudi M, Kovacs S, Alkasassbeh M (2017) Evaluation of machine learning algorithms for intrusion detection. In: 15th International symposium on intelligent systems and informatics, Subotica, Serbia, pp 14–16.

[4] Vinayakumar R, Alazab M, Soman K, Poornachandran P, Al-Nemrat A, Venkatraman S (2019) Deep learning approach for intelligent intrusion detection system. IEEE Access 7:14525–41550.

[5] Butun I, Morgera S, Sankar R (2014) A survey of intrusion detection systems in wireless sensor networks. IEEE Commun Surv Tutorials 16(1):266–282.

[6] Alazab A, Hobbs M, Abawajy J, Khraisat A, Alazab M (2014) Using response action with intelligent intrusion detection and prevention system against web application malware. Inf Manage Comput Secur, 22(5):431–449.

[7] Aburomman, Reaz M,"A survey of intrusion detection systems based on ensemble and hybrid classifiers. Comput Secur 65:135–152.

[8] Buczak, Guven E (2016) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutorials 18(2):1153–1176.

[9] Qassim Q, Zin A, Aziz M (2016) Anomalies classification approach for network-based intrusion detection system. Int J Netw Secur 18(6):1159–1172.

[10] Vimala S, Khanaa V, Nalini C (2019) A study on supervised machine learning algorithm to improvise intrusion detection systems for mobile ad hoc networks. Clust Comput 22:4065–4074

[11] Ahmed M, Mahmood AN, Hu J (2016) A survey of network anomaly detection techniques. J Netw Comput Appl 60:19–31.

[12] Feng W, Zhang Q, Hu G, Huang JX (2014) Mining network data for intrusion detection through combining svms with ant colony networks. Futur Gener Comput Syst 37:127–140.

[13] Li L, Yu Y, Bai S, Hou Y, Chen X (2017) An effective two-step intrusion detection approach based on binary classification and k-NN. IEEE Access 6:12060–12073.

[14] Liu J, He J, Zhang W, Ma T, Tang Z, Niyoyita JP, Gui W (2019) ANID-SEoKELM: adaptive network intrusion detection based on selective ensemble of kernel ELMs with random features. Knowl Based Syst 177:104–116.

[15] Khonde SR, Ulagamuthalvi V (2022) Blockchain: secured solution for signature transfer in distributed intrusion detection system. Comput Syst Sci Eng 40(1):37–51.

[16] Khonde SR, Ulagamuthalvi V (2022) Hybrid intrusion detection system using blockchain framework. Eurasip J Wirel Commun Netw 58.

[17] Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H (2020) Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. J Inf Secur Appl 50:102–419.

[18] Garg S, Kaur K, Batra S, Aujla GS, Morgan G, Kumar N, Zomaya AY, Ranjan R En-abc: an ensemble artificial bee colony based anomaly detection scheme for cloud environment. J Parallel Distrib Comput 135:219–233.

[19] Wu K, Chen Z, Li W (2018) A novel intrusion detection model for a massive network using convolutional neural networks. IEEE Access 6:50850–50859.

[20] Xiao Y, Xing C, Zhang T, Zhao Z (2019) An intrusion detection model based on feature reduction and convolutional neural networks. IEEE Access 7:42210–42219.

[21] R. Chitrakar and H. Chuanhe, ''Anomaly detection using support vector machine classification with k-medoids clustering,'' in Proc. 3rd Asian Himalayas Int. Conf. Internet, Nov. 2012, pp. 1–5.

[22] I. P.-B. A. Syarif and G. Wills, ''Unsupervised clustering approach for network anomaly detection,'' in Proc. Int. Conf. Netw. Digit. Technol., Berlin, Germany, 2012, pp. 135–145.

[23] K. Moh, M. Aung, and N. N. Oo, ''Association rule pattern mining approaches network anomaly detection,'' in Proc. Int. Conf. Future Comput. Technol., Singapore, 2015, pp. 164–170.

[24] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão, and M. L. Proença, ''Network anomaly detection system using genetic algorithm and fuzzy logic,'' Expert Syst. Appl., vol. 92, pp. 390–402, Feb. 2018.

[25] N. T. Pham, E. Foo, S. Suriadi, H. Jeffrey, and H. F. M. Lahza, ''Improving performance of intrusion detection system using ensemble methods and

feature selection,'' in Proc. Australas. Comput. Sci. Week Multiconference, Jan. 2018, pp. 1–6.

[26] I. Sharafaldin, A. Gharib, A. H. Lashkari, and A. A. Ghorbani, ''Towards a reliable intrusion detection benchmark dataset,'' Softw. Netw., vol. 2017, no. 1, pp. 177–200, 2017.

[27] A. M. Al Tobi and I. Duncan, ''KDD 1999 generation faults: A review and analysis,'' J. Cyber Secur. Technol., vol. 2, nos. 3–4, pp. 164–200, Oct. 2018.

[28] N. Moustafa and J. Slay, ''UNSW-NB15: A comprehensive data set for network intrusion detection systems,'' in Proc. Mil. Commun. Inf. Syst., 2015, pp. 1–6.

[29] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, ''Toward generating a new intrusion detection dataset and intrusion traffic characterization,'' in Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy, 2018, pp. 108–116.

[30] Hulk—Packet Storm. Accessed: Aug. 22, 2020. [Online]. Available: https://packetstormsecurity.com/files/112856/HULK-Http-UnbearableLoad-King.html

[31] 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,'' Inf. Secur. J., Global Perspective, vol. 25, nos. 1–3, pp. 18–31, Apr. 2016.

[32] Cyber Kill Chain—Lockheed Martin. Accessed: Aug. 27, 2020. [Online]. Available: https://www.lockheedmartin.com/en-us/capabilities/ cyber/cyber-kill-chain.html

[33] A. Divekar, M. Parekh, V. Savla, R. Mishra, and M. Shirole, ''Benchmarking datasets for anomaly-based network intrusion detection: KDD CUP 99 alternatives,'' in Proc. IEEE 3rd Int. Conf. Comput., Commun. Secur. (ICCCS), Kathmandu, Nepal, Oct. 2018, pp. 1–8.

[34] P. Gil. Cleaning Big Data—Forbes. Accessed: Aug. 26, 2020. [Online]. Available: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-datascience-task-survey-says/#79e15eaa6f63

[35] Documentation—Argus Accessed: Aug. 27, 2020. [Online]. Available: https://openargus.org/documentation,

[36] Online Manual—Tcptrace. Accessed: Aug. 27, 2020. [Online]. Available: http://www.tcptrace.org/manual.html

[37] M. Alkasassbeh, ''An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods,'' J. Theor. Appl. Inf. Technol., vol. 95, no. 22, pp. 5962–5976, 2017.

[38] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, ''Building an intrusion detection system using a filter-based feature selection algorithm,'' IEEE Trans. Comput., vol. 65, no. 10, pp. 2986–2998, Oct. 2016.

[39] CybersecurityUpdate—WebProNews. Accessed: Aug. 21, 2020. [Online]. Available: https://www.webpronews. com/cisco-cybersecurity-threats/

[40] Z. M. Smith, E. Lostri, and J. A. Lewis, ''The hidden costs of cybercrime,'' in Proc. McAfee, 2020, p. 3.