IoT Intrusion Detection Using Feature Selection Method

Subjects: Computer Science, Information Systems

Contributor: Khalid Albulayhi, Qasem Abu Al-Haija, Suliman A. Alsuhibany, Mohammad Ashrafuzzaman, Frederick Sheldon

The Internet of Things (IoT) ecosystem has experienced significant growth in data traffic and consequently high dimensionality. Intrusion Detection Systems (IDSs) are essential self-protective tools against various cyber-attacks. However, IoT IDS systems face significant challenges due to functional and physical diversity. These IoT characteristics make exploiting all features and attributes for IDS self-protection difficult and unrealistic.

Keywords: cybersecurity ; feature selection ; Internet of Things (IoT)

1. Introduction

New cybersecurity risks have emerged owing to organizations deploying Internet of things (IoT) devices in IT (information technology) and OT (operational technology) environments. Such new risks threaten to undermine structural tenets such as safety, mobility, efficiency, and security of operational ecosystems. New threat vectors not only affect technological aspects of the lives but also pose a risk towards financial and physical wellbeing [1]. The threat of attack has brought several insecurities to online privacy, social networks, business, and critical infrastructure [2][3]. Therefore, the development of resilient strategies has become an essential part of dynamical environments such as the IoT ecosystem. IoT is a constantly evolving emerging technology set [4][5] that changes the security and risk schematics of automated networked systems [6][7]. IoT has spread into a wide range of human systems to shape the core of our industrial society to become the man–machine interface of life. By 2024, IoT is expected to reach 83 billion devices operationally [8]. IoT applications include smart cities, smart homes, and intelligent transportation. These applications deploy IoT devices to increase productivity and reduce costs by using ’plug-n-play’ kits that do not require extensive prior device knowledge. Such a ‘plug-n-play’ configuration increases the risk of cyber-misbehavior. Compounding factors include the typical mixture of multitudes of wired or wireless communications that employ cloud-connected embedded systems used by consumers to interconnect with each other [9][10].

Intrusion detection systems (IDS) are widely used to improve security posture in an IT infrastructure. An IDS is considered a suitable and practical approach to detect attacks and assure network security by safeguarding against intrusive hackers [3]. Anomaly-based IDS approaches can efficiently detect zero-day (unknown) attacks [11][12]. An intrusion can be defined as a sequence of unexpected activities locally or globally, harming network confidentiality, integrity, and/or availability (i.e., the CIA triad) [13][14]. The network traffic consists of packets associated with packet header fields. Features related to those instances are important to define the purpose of detecting anomalies. The purpose of an IDS is to detect and/or prevent abnormal misbehavior (i.e., unauthorized use), both passive and active network intruder activities, and thus improve CIA.

In recent times, machine learning (ML)-based approaches have been employed for intrusion detection in IoTs IDSs [3][15][16][17][18][19][20][21]. Existing IDSs assume that the IoT devices have the same feature pattern and packet types. However, IoT devices vary in some respects, such as hardware characteristics and functionality, computational capability, and different abilities for generating various features [11][22]. The features become sparse when nodes are aggregated to create data, and the irrelevant features (attributes) are set to either nulls or zeros. Data sparsity is one of the disadvantages that affect the accuracy and efficiency of data modeling. Feature selection, an important part of a machine learning-based solution, plays an important role in increasing detection accuracy and speed of the training phase. Several feature selection techniques have been proposed to improve detection of anomalous behavior variants such as Flexible Mutual Information-based Feature Selection (FMIFS) [23], Modified Mutual Information-Based Feature Selection (MMIFS) with Support Vector Machine (SVM) [24], and SVM with Neural Networks (NN) [25]. Those approaches/models and other recent state-of-the-art studies have been presented in the related work section. *Detection accuracy of anomaly-based IDSs is considered the main challenge in the IoT ecosystem due to the constantly evolving nature of the IoT environment* [26][27].

2. Feature Selection

IoT datasets are of intrinsically high dimensionality represented by *n* instances and *m* columns (features) [11]. The data matrix is $X \in \mathbb{R}^{N \times M,}$ and the Y is the target variable(s) (class(es)). A target instance (class) may be either discrete or continuous, and the model can also be dynamic or static. A feature selection (FS) enhances model performance by reducing dimensionality. FS can be defined as a subset of $P \ll M$ features, i.e., $X_{FS} \in \mathbb{R}^{N \times P}$, where p are relevant features of the target class.

Feature selection endeavors to eliminate irrelevant and redundant features and to choose the most pertinent and important features. Furthermore, the FS process usually improves the general performance and data dimensionality, reducing the cost of classification and prediction by reducing the time complexity for building the model. On the other hand, applying all features in the IDS model includes several drawbacks: (i) the computational overhead is increased, and training and testing time are slower, (ii) storage requirements increase due to the large number of features, (iii) the error rate of the model increases because irrelevant features diminish the discriminating power of the relevant features as well as reduce accuracy. FS approaches can be characterized into five categories: (i) filter-based, (ii) wrapper-based, (iii) embedded-based, (vi) hybrid-based, and (v) learning-based. The filter method gives weights to each feature (i.e., dimension), sorts them based on these weights, and then uses those subsets of features to train the model for either classification or prediction. Therefore, the process of feature selection is independent of the classification/prediction techniques. Numerous statistical measures are used in filtering methods to obtain feature subsets.

The model, using a particular FS method, initially uses all features but subsequently omits unrelated features to address the *curse of dimensionality* problem. This refining is designed to acquire the best subset of features based on statistical gauges such as information gain (IG) and gain ratio (GR), Pearson’s correlation (PC) [28], chi-square (Chi12) [29], and mutual information (MI) [30][31][32][33]. The wrapper method is considered a black box technique [34]. Inductive algorithms are used to select feature subsets in the wrapper method, whereas filter methods are independent of the inductive algorithm. In addition, wrapper methods are more complex and expensive computationally than filter methods because they rely on iterating the learning systems (i.e., ML-derived models) several times until a subset of relevant features is reached. Moreover, the wrapper method accounts for the influence of the model performance on the feature subsets and strives to achieve high classification accuracy.

Embedded methods are incorporated with ML algorithms to select a feature subset during the learning process. The blending of feature selection approaches is used during the learning process to achieve advantages by improving classification, accuracy, and computational cost. Embedded methods can avoid retraining the model when the model needs to add a new feature to the subset. Concerning the structure of the embedded approach, the feature selection process is integrated with the classification algorithm and simultaneously performs feature selection such as random forest, LASSO (Least Absolute Shrinkage and Selection Operator), and L1 regularization [35]. Embedded methods are computationally less intensive than wrapper methods. However, they still have high computational complexity. Furthermore, the selected feature subset result depends on the chosen learning algorithm. Thus, embedded methods endeavor to find the best feature subset during model building by selecting each feature individually. Furthermore, they derive significant advantages in terms of model interaction, accuracy, fewer variables, and computational cost than previous approaches.

Information Gain (IG) [36] is one of the most widely used approaches in preparing features from a filter-based approach. That is, IG provides a classification ranking of all attributes (features) related to the target (class). Then a threshold is assigned to select several features according to the order obtained. Accordingly, a feature that strongly correlates with the target is considered a relevant feature and irrelevant (or redundant) otherwise. However, a weakness of the IG criterion is a bias favoring features with more values, especially when they are not more informative. Thus, IG between the feature in *X* and the variable (target) *y* is given here in Equation (1):

$$\mathrm{IG} = H(Y) - H(Y|X) = H(X) - H(X|Y) \tag{1}$$

where $H(x)$ is the entropy of x given y. The entropy of y is defined by Equation (2):

$$H(y) = -\sum_{y \in Y} p(y) \log_2 (p(y)) \tag{2}$$

where $p(y)$ is the marginal probability of y on all values of Y. Note, Y is a finite set. Moreover, the conditional entropy of Y given the random variable X is shown in Equation (3):

$$H(Y|X) = -\sum_{x\in X} p(x) \sum_{y\in Y} (p(Y|X) \log_2 (p(Y|X)) \quad (3)$$

where $p(y|x)$ is the conditional probability of y given x.

IG is a symmetrical measure such as IG (x, y) = IG (y, x), as shown in Equation (1).

The information gained about Y after observing X is equal to the information gained about X after observing Y.

Gain Ratio (GR) [37] is the non-symmetrical measure introduced to compensate for the bias of the IG attribute evaluation. The GR formula is given in Equation (4):

$$\mathrm{GR} = \mathrm{IG} / (H(X)) \quad (4)$$

3. IoT Intrusion Detection Using Feature Selection Method

Significant and fruitful efforts have endeavored to address the security concerns of recent years for the IoT ecosystem. Several new IoT security technologies were established by pairing artificial intelligence techniques and cybersecurity virtues. Several promising state-of-the-art studies have been conducted for IoT security using machine learning (ML) and deep learning (DL) techniques [38][39][40][41][42][43][44][45][46][47]. However, only a few were developed by investigating the impact of using different feature selection approaches to improve prediction and classification accuracy. For instance, Albulayhi et al. [11] have proposed and implemented a new minimized redundancy discriminative feature selection (MRD-FS) technique to resolve the issue of redundant features. The discriminating features have been selected based on two criteria, i.e., representativeness and redundancy. Their model was evaluated utilizing the BoT-IoT dataset. Ambusaidi et al. [23] presented a flexible, mutual information-based feature selection technique (FMIFS) that chooses the best features to enhance the classification algorithm. The proposed model was evaluated using three datasets (NSL-KDD, KDD Cup 99, and Kyoto 2006). The Least Square Support Vector Machine-based IDS (LSSVM-IDS) was used to measure performance. Ambusaidi et al. [23] showed 99.79% accuracy, 99.46% detection rate (DR), and 0.13% FPR over the KDD99 dataset. However, their employed datasets are not up-to-date (date back to 2009, 1999, and 2006 for NSL-KDD, KDD-Cup99, and Kyoto datasets, respectively) and do not fully represent the IoT cyberattacks.

Similarly, Amiri et al. [24] proposed a modified mutual information-based feature selection technique (MMIFS) applied with the SVM to improve the accuracy performance of the classification and to (highly) efficiently detect the various attack types. They demonstrated how high data dimensionality could be enhanced using the feature selection technique. Note, high dimensionality, even if applied to a high-quality ML approach, produces poor detection rate and accuracy performance. MMIFS can reduce features to only eight features (out of 41). For instance, MMIFS with SVM using only eight features, and DR achieved 86.46%. In the first phase, data normalization and reduction are applied by dividing every attribute (feature) value by its maximum value. In the next phase, feature selection is applied based on the imported training data. Further, MMIFS initially takes the feature set as the empty set. In more detail, it calculates the mutual information of the features concerning the class target and then picks the first feature with the maximum mutual information value.

References

1. Turton, W.; Mehrotra, K. Hackers breached colonial pipeline using compromised password. Available online: https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password (accessed on 1 October 2021).
2. Smadi, A.A.; Ajao, B.T.; Johnson, B.K.; Lei, H.; Chakhchoukh, Y.; Al-Haija, Q.A. A Comprehensive Survey on Cyber-Physical Smart Grid Testbed Architectures: Requirements and Challenges. Electronics 2021, 10, 1043.
3. Albulayhi, K.; Smadi, A.A.; Sheldon, F.T.; Abercrombie, R.K. IoT Intrusion Detection Taxonomy, Reference Architecture, and Analyses. Sensors 2021, 21, 6432.

4. Farooq, M.U.; Waseem, M.; Mazhar, S.; Khairi, A.; Kamal, T. A review on internet of things (IoT). Int. J. Comput. Appl. 2015, 113, 1–7.
5. Aborujilah, A.; Nassr, R.M.; Al-Hadhrami, T.; Husen, M.N.; Ali, N.A.; Al-Othmani, A.; Syahela, N.; Ochiai, H. Security Assessment Model to Analysis DOS Attacks in WSN. In International Conference of Reliable Information and Communication Technology; Springer: Berlin/Heidelberg, Germany, 2019.
6. Agrawal, K.; Kamboj, N. Smart agriculture using IOT: A futuristic approach. Int. J. Inf. Dissem.Technol. 2019, 9, 186–190.
7. Pawar, P.; Trivedi, A. Device-to-device communication based IoT system: Benefits and challenges. IETE Tech. Rev. 2019, 36, 362–374.
8. CISOMAG. IoT Connections to Reach 83 Billion by 2024: Report. Available online: https://cisomag.eccouncil.org/iot-connections-to-reach-83-billion-by-2024-report/ (accessed on 12 July 2021).
9. Kumar, S.; Solanki, V.K.; Choudhary, S.K.; Selamat, A.; González Crespo, R. Comparative Study on Ant Colony Optimization (ACO) and K-Means Clustering Approaches for Jobs Scheduling and Energy Optimization Model in Internet of Things (IoT). Int. J. Interact. Multimed. Artif. Intell. 2020, 6, 107–116.
10. Nimbalkar, P.; Kshirsagar, D. Feature selection for intrusion detection system in Internet-of-Things (IoT). ICT Express 2021, 7, 177–181.
11. Albulayhi, K.; Sheldon, F.T. An Adaptive Deep-Ensemble Anomaly-Based Intrusion Detection System for the Internet of Things. In Proceedings of the 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 10–13 May 2021.
12. Eskandari, M.; Janjua, Z.H.; Vecchio, M.; Antonelli, F. Passban IDS: An Intelligent Anomaly-Based Intrusion Detection System for IoT Edge Devices. IEEE Internet Things J. 2020, 7, 6882–6897.
13. Abraham, A.; Grosan, C.; Martin-Vide, C. Evolutionary design of intrusion detection programs. Int. J. Netw. Secur. 2007, 4, 328–339.
14. Ilgun, K.; Ustat, A. A Real-Time Intrusion Detection System for Unix. Master’s Thesis, University of California Santa Barbara, Santa Barbara, CA, USA, 1992.
15. Verma, A.; Ranga, V. Machine learning based intrusion detection systems for IoT applications. Wirel. Pers. Commun. 2020, 111, 2287–2310.
16. Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput. Sci. 2021, 2, 1–20.
17. Siddiqi, M.A.; Pak, W. An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection. IEEE Access 2021, 9, 137494–137513.
18. Latif, S.; Zou, Z.; Idrees, Z.; Ahmad, J. A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network. IEEE Access 2020, 8, 89337–89350.
19. Heigl, M.; Weigelt, E.; Fiala, D.; Schramm, M. Unsupervised Feature Selection for Outlier Detection on Streaming Data to Enhance Network Security. Appl. Sci. 2021, 11, 12073.
20. Sarker, I.H. Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective. SN Comput. Sci. 2021, 2, 1–16.
21. Balogh, S.; Gallo, O.; Ploszek, R.; Špaček, P.; Zajac, P. IoT Security Challenges: Cloud and Blockchain, Postquantum Cryptography, and Evolutionary Techniques. Electronics 2021, 10, 2647.
22. Alrubayyi, H.; Goteng, G.; Jaber, M.; Kelly, J. Challenges of Malware Detection in the IoT and a Review of Artificial Immune System Approaches. J. Sens. Actuator Networks 2021, 10, 61.
23. Ambusaidi, M.A.; He, X.; Nanda, P.; Tan, Z. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. IEEE Trans. Comput. 2016, 65, 2986–2998.
24. Amiri, F.; Yousefi, M.R.; Lucas, C.; Shakery, A.; Yazdani, N. Mutual information-based feature selection for intrusion detection systems. J. Netw. Comput. Appl. 2011, 34, 1184–1199.
25. Sung, A.H.; Mukkamala, S. Identifying important features for intrusion detection using support vector machines and neural networks. In Proceedings of the 2003 Symposium on Applications and the Internet, Orlando, FL, USA, 27–31 January 2003.
26. Jose, S.; Malathi, D.; Reddy, B.; Jayaseeli, D. A Survey on anomaly-based host intrusion detection system. In Journal of Physics: Conference Series; IOP Publishing: Bristol, UK, 2018.
27. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity 2019, 2, 1–22.

28. Biesiada, J.; Duch, W. Feature Selection for High-Dimensional Data—A Pearson Redundancy Based Filter, in Computer Recognition Systems 2; Springer: Berlin/Heidelberg, Germany, 2007; pp. 242–249.
29. Jin, X.; Xu, A.; Bie, R.; Guo, P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. In International Workshop on Data Mining for Biomedical Application; Springer: Berlin/Heidelberg, Germany, 2006.
30. Thang, N.D.; Lee, Y.-K. An improved maximum relevance and minimum redundancy feature selection algorithm based on normalized mutual information. In Proceedings of the 2010 10th IEEE/IPSJ International Symposium on Applications and the Internet, Seoul, Korea, 19–23 July 2010.
31. Estévez, P.A.; Tesmer, M.; Perez, C.A.; Zurada, J.M. Normalized Mutual Information Feature Selection. IEEE Trans. Neural Networks 2009, 20, 189–201.
32. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 1226–1238.
33. Kwak, N.; Choi, C.-H. Input feature selection by mutual information based on Parzen window. IEEE Trans. Pattern Anal. Mach. Intell. 2002, 24, 1667–1671.
34. Kohavi, R.; John, G.H. Wrappers for feature subset selection. Artif. Intell. 1997, 97, 273–324.
35. Osman, H.; Ghafari, M.; Nierstrasz, O. Automatic feature selection by regularization to improve bug prediction accuracy. In Proceedings of the 2017 IEEE Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE), Klagenfurt, Austria, 21 February 2017.
36. Quinlan, J.R. Induction of decision trees. Mach. Learn. 1986, 1, 81–106.
37. Han, J.; Pei, J.; Kamber, M. Data Mining: Concepts and Techniques; Elsevier: Amsterdam, The Netherlands, 2011.
38. Abu Al-Haija, Q.; Zein-Sabatto, S. An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks. Electronics 2020, 9, 2152.
39. Bendiab, G.; Shiaeles, S.; Alruban, A.; Kolokotronis, N. IoT Malware Network Traffic Classification using Visual Representation and Deep Learning. In Proceedings of the 2020 6th IEEE Conference on Network Softwarization (NetSoft), Ghent, Belgium, 29–31 July 2020.
40. AAbu Al-Haija, Q.; McCurry, C.D.; Zein-Sabatto, S. Intelligent Self-reliant Cyber-Attacks Detection and Classification System for IoT Communication Using Deep Convolutional Neural Network. In Selected Papers from the 12th International Networking Conference, Rhodes, Greece, 19–21 September 2020; Springer: Cham, Switzerland, 2021.
41. Taher, K.A.; Jisan, B.M.Y.; Rahman, M. Network intrusion detection using supervised machine learning technique with feature selection. In Proceedings of the 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 10–12 January 2019.
42. Gao, X.; Shan, C.; Hu, C.; Niu, Z.; Liu, Z. An Adaptive Ensemble Machine Learning Model for Intrusion Detection. IEEE Access 2019, 7, 82512–82521.
43. Sapre, S.; Ahmadi, P.; Islam, K. A robust comparison of the KDDCup99 and NSL-KDD IoT network intrusion detection datasets through various machine learning algorithms. arXiv 2019, arXiv:1912.13204.
44. Chowdhury MM, U.; Hammond, F.; Konowicz, G.; Xin, C.; Wu, H.; Li, J. A few-shot deep learning approach for improved intrusion detection. In Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017.
45. Javaid, A.; Niyaz, Q.; Sun, W.; Alam, M. A Deep Learning Approach for Network Intrusion Detection System. In In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), New York, NY, USA, 3–5 December 2015.
46. Imamverdiyev, Y.; Sukhostat, L. Anomaly detection in network traffic using extreme learning machine. In Proceedings of the 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 12–14 October 2016.
47. Al-Haijaa, Q.A.; Ishtaiwia, A. Machine Learning Based Model to Identify Firewall Decisions to Improve Cyber-Defense. Int. J. Adv.Sci. Eng. Inf. Technol. 2021, 11, 1688–1695.

Retrieved from https://encyclopedia.pub/entry/history/show/58257