IoT Based Smart Cities Application for Detecting Cyber-Attacks Using Machine Learning

Zohaib Ahmad[1], Arslan Munir[2,] Arslan Aslam[3]

[1]University of Engineering and Technology (UET), Lahore, 53000, Pakistan
[2]Bahauddin Zakeria University (BZU), Multan, 61000, Pakistan
[3]The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan
[*]Corresponding Author: Aleeza Nouman. Email: zohaib.ahamd@kics.edu.pk

Abstract: In recent years, the increased adoption of Internet of Things (IoT) apps has led to the emergence of smart cities. Smart cities are designed to improve the value of public services and the welfare of its citizens. To do this, they make use of IoT-enabled technology, communication, and applications. However, with increasing smart city networks, there is a greater chance of online security risks and attacks. IoT devices are especially vulnerable to such dangers and malicious attacks due to their connection with sensors and cloud servers. That's why it is essential to develop measures to avoid such assaults and prevent IoT gadgets from breakdown.

This research proposes an IoT-based model for smart city applications to detect cyber-attacks by using machine learning. This proposed model performs better than the Deep Neural Network Expert System (DeNNeS) approach and provides more reliable and secure functionality to the power networks (Smart Traffic, Smart Grid & Smart Buildings) against cyber-attacks by using machine learning in the future.

Keywords: Smart cities; Internet of Things (IoT) ; cyber attacks; machine learning

1 Introduction

The IoT is a network of interlinked systems that encourages easy information sharing among gadgets, such as intelligent house sensors, green sensors, cars as well as roadside sensors, medical devices, manufacturing robots, as well as surveillance equipment. Along with the amount of connected IoT devices hitting 27 billion in 2017 as well as being predicted to reach roughly 125 billion in 2030, the IoT's usage in communities and services has recently seen a considerable surge. Different services, technologies, and protocols are used by IoT devices. Future IoT infrastructure maintenance will become extremely difficult as a result, which ultimately exposes the system to unwanted vulnerability [1].

Internet of Things (IoT) devices remain used in clever city applications, so cyberattacks can gain unauthorized admittance to information about citizens' daily activities without the worker's or manager's knowledge or reconfigure strategies to an unsecured setting. The number of assaults on the IoT platform in 2019 that attempted to influence the connected nature of such devices increased by 600%, according to Symantec [2-3].

Applications for smart cities provide many security difficulties. First, zero-day attacks can happen by taking advantage of flaws in various protocols used by applications for smart cities. Second, can network-based cyberattacks be intelligently detected before they impair the functioning of smart cities? Third, the IoT machines utilized in intelligent cities have restricted aboard capacity for protection procedures, are

often resource-constrained (e.g., memory-wise), and transport collected information to cloud attendants for managing. Current intrusion detection systems (IDS) do not account for IoT machines. Is it possible to create an IDS specifically for IoT networks by combining all these issues?

The cloud computation ecosystem, which gives gradually more sophisticated CPUs as well as sufficient memorial resources, is where the data gathered after the IoT system is collected. However, along with the latest growth in IoT devices, the amount of data being carried after the IoT station layer to the cloud has expanded quickly, which causes delays as well as blockage issues in the cloud. Fog computing is intended as a potential remedy for many issues. The computing burden that was initially sent to the cloud can be shared more widely across the fog layer devices. This fixes the issue with data storage and transfers while reducing energy use, network traffic, and latency. Additionally, it intends to move processing closer to edge devices so that IoT-based smart city applications may be responded to quickly. Cyberattack recognition in the cloud layer has two advantages. First off, if attacks are discovered early in the fog layer, the ISP or system manager can take the appropriate precautions to avert significant harm. Second, it won't interfere with daily living as usual in cities. To address the problem, certain solutions (such as signature basis techniques) have been put forth in the literature. A compilation of earlier created names (attacks) is compared to the most recent suspicious samples in the signature-based technique. Assaults may go undetected or cause false alarms if the name removal approach is not entirely able to obtain the distinctive features of attacks or damage people. This method has a substantial processing overhead and is unsuitable for identifying unidentified attacks. Machine learning algorithms require less processing time than other techniques and can identify assaults as they happen [4].

This research investigates that an assault and anomaly identification method depend upon ML for IoT-based smart city apps. This method can recognize contaminated Internet of Things (IoT) gadgets, a cloud services difficulty. The technology uses a training model on distributed fog networks (FN) to identify assaults and anomalies and learn from IoT devices. Researchers are driven to create an ensemble model of classifiers since it is frequently insufficient to use a single classifier to create an effective IDS. Ensemble approaches combine multiple models to get a single final model by considering a wide range of models. According to research, the ensemble model outperforms the single classifier in terms of performance. To guarantee improved performance by the combination technique, several variables (such as feature selection and base classifier) must be carefully considered. The trio of bagging, boosting, and stacking ensemble approaches are the most effective. In this study, we employ both individual classifiers and ensemble methods to improve IDS accuracy, precision, recall, as well as F1-Score. We study machine learning-based attacks as well as anomaly detection in disseminated FN over IoT-based methods which is how the paper's contributions are summed up. To identify assaults and anomalies, existing efforts have typically used signature-based approaches. These methods have large overhead costs and are susceptible to known dangers. This research investigates the viability of ensemble-based learning in this study, Int. J. Environ. Res. Public Health 2020, 17, 9347 3 of 21, equated to one model classifier for detecting cyberattacks in IoT-based intelligent city apps. In addition, we compare a multi-class classification setup to the second-class calculation used in most pertinent publications. In addition to these, we consider the addition of quality choice as well as cross authentication since many conventional machine learning techniques have not remained adequately addressed in the body of literature already available for this field. An extensive study that considers the integrations mentioned above demonstrates that an ensemble of machine learning-based classifiers performs well than a single classifier at correctly classifying assaults and their categories [5].

2 Literature Review

Many Several research work have been issued in the literature to enhance IDS accuracy. This portion highlights current significant efforts which have integrated ensemble processes and ML approaches. Pahl, as well as Aubet, suggested a ML-based technique to anticipate IoT maintenance conduct in a dispersed multi-dimensional IoT site. This method employs K-means and BIRCH-based clustering methods inside an IoT site to continuously learn microservice models. In this instance, the cluster centers are consolidated into one if they are within the triplet times traditional deviation space. The model updates cluster formation using a communication paradigm for online learning. This technique has a 96.5 percent overall precision for anomaly detection and a 0.2 percent false positive rate [6].

To identify On as well as Off attacks in a manufacturing IoT location coming after hostile system points, a combined trust light probe-based protection method remained created. In this case, an IoT network could be targeted by a malicious node when it is in an On or Off state thanks to the "On and Off attack." With the measurement of confidence approximation for every neighbor point, the structure was created for the detection of abnormalities using a light analysis routing method. Making use of the NSL-KDD wide open-resource dataset, which describes damage data in the spread as well as centralized systems, Diro, as well as Chilamkurti, suggested a deep-rooted knowledge model identify spread attacks in a community IoT system. They then matched the performance of the thick model including a superficial NN. By using of dual-class (regular and attack) as well as four-class (normal, categories, they assessed the accuracy of the deep and shallow models. Their model attained accuracies of 99.2 percent and 98.27 percent for binary-class and multi-class identification and 95.22 percent as well as 96.75 percent, respectively, for the deep as well as superficial models [7].

Since of the negative impacts of low point rate attacks such client to root as well as remote-to-local (R2L) incidents, Pajaud et al. presented a double-stage height decrease as well as a classification method to identify irregularity in IoT backbone systems. They minimized the dataset's properties using principal component analysis (PCA) as well as linear discriminate analysis (LDA), then utilized naive Bayes as well as K-nearest Neighbor (KNN) to locate abnormality, achieving an identification rate of 84.82 percent. Extreme learning machine (ELM) method-based attack detection was first introduced by Kozik et al. Data acquired from the fog computing environment in NetFlow format can be computed and effectively analyzed thanks to ELM's architecture and features. These three cases—scanning, command and control, and infected host—were the focus of this work, which achieved accuracy levels of 99 percent, 76 percent, and 95 percent, respectively [8]

An approach based on data analysis that circumvents the data processing burden of signature-based techniques was presented by Hasan et al. to identify assaults on IoT infrastructure. When the system exhibits any unusual activity, its suggested approach can detect it and shield the system from attacks. They used the openly accessible IoT dataset for their experiment. They investigated several machine learning methods, including DT, RF, LR, SVM, as well as ANN, with the RF classifier producing the greatest outcomes. A random forest-based abnormality detection approach that may find contaminated IoT machines at dispersed mist points was proposed. Only 12 of the dataset's 49 attributes were considered by their binary random forest (RF) classifier when testing it on the UNSW-NB15 dataset. Extra Tree Classifier was used to retrieve these 12 characteristics. Analysis of their performance revealed that they had a 0.02 percent false positive rate and 99.34 percent accuracy. On the NSL-KDD, UNSW-NB15, WSN-DS, as well as CICIDS 2017 datasets, a deep-seated learning model was investigated to recognize cyberattacks. They determined that the DL model outperforms the other machine learning methods [9].

Multiple ensemble methods-based IDSs are suggested in the literature to improve accuracy compared to basic classifiers. Gain ratio (GR) feature choice methodology was applied in ANN and Bayesian net-based ensemble method, and performance was assessed on KDD'99 and S-KDD datasets, where ensemble

techniques obtained 99.42 percent and 98.07 percent accuracy, respectively. An ensemble technique that incorporates Naive Bayes, Bayesian Net, as well a decision tree classifier was proposed by Haq et al. in their study. Using feature selection methods such as Best First Search, Genetic, and Rank Search, they were able to extract the common features. When examined using a 10-fold cross-validation procedure, the ensemble technique obtained a true positive rate of 98%. Restreet was utilized as the base classifier in the bagging ensemble method that Gaikwad et al. proposed. On the NSL-KDD dataset, their model's accuracy was 81.29 percent. Using alternating decision trees (AD Tree) and KNN, Jabbar et al. introduced an ensemble method and performance testing showed that it outperformed previous methods in terms of detection rate (99.8%) [10].

An integrated framework known as a "smart city" combines IoT, intelligent structures, as well as information as well as communication technology (ICT) to recover the excellence as well as the effectiveness of various city facilities like transport, healthiness schemes, contamination management, as well as power delivery. IoT products and networks are sensitive to attacks for several causes. First off, the mainstream of IoT gadgets suffer from poor managing capability due to their reduced supplies (e.g., low managing energy as well as recall). The second component contributing to lag in cloud information centers is the interconnectedness of IoT gadgets utilizing multiple protocols, as well as the enhancing number of IoT gadgets. Third, IoT gadges can occasionally be left unattended, making it possible for a burglar to get physical access to them. Fourth, most data transmissions are wireless, making them vulnerable to spying. As a result, traditional IDS structures frequently fall short of precisely identifying the IoT threat. To connect to smart city routers as well as machines at numerous locations, including houses, retail malls, restaurants, hotels, and airfields, an assailant can successfully hack susceptible IoT devices. By doing this, an assailant who takes control of these IoT devices may get precise data, including credit card information, streaming video, as well as other types of personal data [11].

Most of the techniques have been utilized while using and constructing many smart as well as intelligent structures like ML approaches [12-15], ,cloud computing [16-17], MapReduce [18], Explainable Artificial Intelligence [19], mining techniques [20] and different security based processes [21] that can provide help in designing developing solutions for the growing tasks in designing smart cloud-based observing management systems.

3 Proposed Methodology

Smart cities are rapidly adapting to the Internet of Things (IoT) technologies in order to improve the quality of life of its citizens, as well as its operational efficiency. However, with the increased usage of these technologies, the threat of cyber-attacks and other related incidents is ever-growing. To protect against such threats, it is essential to create strategies to prevent and protect IoT systems in the event of a cyber-attack.

research proposal introduces a cyber-attack detection model to detect cyber-attacks in real-time and protect smart city IoT infrastructure. This model will be able to detect cyber-attacks by analyzing network traffic and various factors, such as source and destination IP addresses, ports, and data size. It will also be able to differentiate between benign and malicious activities. Moreover, this model will be able to identify malicious traffic, detect anomalies, and alert the responsible authorities for them to take preventive measures. As a result, this model will help provide secure services to the citizens and ensure the security of the smart city IoT infrastructure. To further strengthen the protection against cyber-attacks, the model should be supplemented with other security measures, such as encryption, secure authentication protocols, and more robust firewalls. Furthermore, it should be tested thoroughly with multiple datasets to ensure accuracy and reliability. Lastly, the proposed model should be continuously monitored and updated to keep up with the ever-evolving cyber-attack landscape. With these measures in place, smart cities will be better equipped to ward off cyber-attacks and protect its IoT infrastructure.

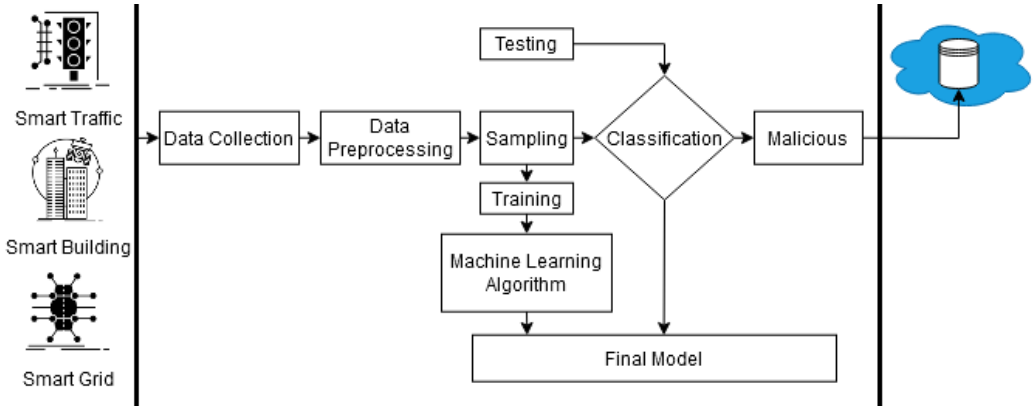

Figure 1: Proposed model

Figure 1 described that the proposed cyber-attack detection model is consisting of the training and the validation phase. The training phase is about data collection from the smart cities' input factors and forwarding these values for the proceeding purpose. In the preprocessing procedure, the noise is removed by using multiple techniques. After the preprocessing, the preprocessed data is sent for sampling. Data sampling is a statistical analysis technique used to select, manipulate, and analyze a representative subset of data points to identify patterns and trends in the larger data set being examined. After the sampling procedure, the data is sent for classification as well as for training through the machine learning algorithm in order to predict whether the malicious or cyber-attack is found or not, and the resultant outcome is stored in the cloud for further proceedings.

4 Limitations & Future Recommendations

As the network infrastructure of smart cities becomes more complex, it is imperative that secure solutions are developed to protect citizens against various cyber-attacks. Machine learning (ML) algorithms are a promising area of research for providing such solutions. ML algorithms can be used to identify malicious behavior from large datasets of IoT sensor data, thus enabling early detection of cyber-attacks. By combining ML with other techniques such as data fusion, pattern recognition and anomaly detection, a more robust and efficient system can be developed for detecting cyber-attacks. Furthermore, ML models can also be trained to detect unusual behavior or anomalous patterns in data collected from smart city sensors, thus allowing for quick identification of cyber-attacks. Additionally, ML models can also be used to identify complex dependencies between various components of the smart city network, allowing for the design of more secure and resilient systems. For example [37], predictive models can be trained using historical data to predict when and where cyber-attacks might take place, thus allowing for better security measures to be taken in advance. Moreover, ML algorithms can also help detect malicious behavior and vulnerabilities in the smart city network, thus allowing for timely remediation and prevention of cyber-attacks.

Finally, ML models can also be used to develop secure authentication protocols for users of smart city networks. By utilizing user authentication data, ML algorithms can be trained to detect malicious activities or unauthorized users. Furthermore, ML models can be used to identify patterns in user behavior and usage patterns, thus providing more secure authentication protocols and reducing the risk of cyber-attacks. In short, ML algorithms can be used to improve the security of smart city networks and provide secure authentication protocols for users.

5 Conclusion

In this paper, the cost-effectiveness of IOT-based smart cities applications is further enhanced by their ability to reduce the burden on existing security protocols. By providing an additional layer of security, these applications can help to ensure that an organization's data and infrastructure are protected against any malicious intent. Additionally, such applications can also help to identify any potential security flaws in the system, allowing the organization to take corrective measures before a cyber-attack takes place. The implementation of IOT-based Smart Cities applications also allows for better coordination between different stakeholders, as they can provide a platform for enhanced communication. This helps to ensure that all stakeholders are on the same page, allowing them to work together towards a common goal. Moreover, such applications can also help to reduce the costs associated with data collection and analysis, as they can automate certain processes and reduce the need for manual labor. Overall, IOT-based Smart Cities applications presents a promising and cost-effective solution to detect and mitigate cyber-attacks. Such applications can also help to optimize the overall efficiency and resilience of a given system, while also providing a platform for improved communication and coordination between different stakeholders. Therefore, organizations should seriously consider the implementation of such applications to improve their security and overall operational performance.

6 References

[1] Roy, P.K., Chowdhary, S.S., Bhatia, R., 2020. A Machine Learning approach for automation of Resume Recommendation system. Procedia Comput. Sci. 167, 2318–2327.

[2] Berrar, D., 2018. Cross-validation. Encycl. Bioinforma. Comput. Biol. ABC Bioinforma. 1–3, 542–545.

[3] Velmurugadass, P., Dhanasekaran, S., Shasi Anand, S., Vasudevan, V., 2020. Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm. Mater. Today Proc. 37, 2653–2659.

[4] Abeshu, A., Chilamkurti, N., 2018. Deep Learning: The Frontier for Distributed Attack Detection in Fog-To-Things Computing. IEEE Commun. Mag. 56, 169–175.

[5] Franceschelli, M., Pilloni, A., Gaspam, A., 2018. A Heuristic approach for Online Distributed optimization of Multi-Agent Networks of Smart Sockets and Thermostatically Controlled Loads based on Dynamic Average Consensus. 2018 Eur. Control Conf. ECC 2018 2541–2548.

[6] Freund, Y., Schapire, R.E., Hill, M., 1996. Experiments with a New Boosting Algorithm Rooms f 2B-428 , 2A-424 g.

[7] Liu, X., Liu, Y., Liu, A., Yang, L.T., 2018. Defending ON-OFF attacks using light probing messages in smart sensors for industrial communication systems. IEEE Trans. Ind. Informatics 14, 3801–3811.

[8] Hasan, M., Islam, M.M., Zarif, M.I.I., Hashem, M.M.A., 2019. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things (Netherlands) 7, 100059.

[9] Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A., 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSP 2018 - Proc. 4th Int. Conf. Inf. Syst. Secur. Priv. 2018-Janua, 108–116.

[10] Zhou, Y., Cheng, G., Jiang, S., Dai, M., 2020. Building an efficient intrusion detection system based on feature selection and ensemble classifier. Comput. Networks 174.

[11] Chakraborty, S., Aich, S., Kim, H.C., 2019. A Secure Healthcare System Design Framework using Blockchain Technology. Int. Conf. Adv. Commun. Technol. ICACT 2019-Febru, 260–264.

[12] Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M. A., Issa, G.F., Ahmad, M., Khan, Muhammad Adnan, 2021. An iomt-enabled smart healthcare model to monitor elderly people using machine learning technique. Comput. Intell. Neurosci. 2021.

[13] Ghazal, T.M., Noreen, S., Said, R.A., Khan, M.A., Siddiqui, S.Y., Abbas, S., Aftab, S., Ahmad, M., 2022. Energy demand forecasting using fused machine learning approaches. Intell. Autom. Soft Comput. 31, 539–553.

[14] Saleem, M., Abbas, S., Ghazal, T.M., Adnan Khan, M., Sahawneh, N., Ahmad, M., 2022. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egypt. Informatics J.
[15] Muneer, S., Raza, H., 2022. An IoMT enabled smart healthcare model to monitor elderly people using Explainable Artificial Intelligence (EAI) 1, 16–22.
[16] Ubaid, M., Arfa, U., Muhammad, H., Muhammad, A., Farooq, S., Saleem, M., 2022. Intelligent Intrusion Detection System for Apache Web Server Empowered with Machine Learning Approaches 1.
[17] Khan, Z., 2022. Used Car Price Evaluation using three Different Variants of Linear Regression 1, 40–49.
[18] Asif, M., Abbas, S., Khan, M. A., Fatima, A., Khan, Muhammad Adnan, Lee, S.W., 2021. MapReduce based intelligent model for intrusion detection using machine learning technique. J. King Saud Univ. - Comput. Inf. Sci.
[19] Muneer, S., Rasool, M.A., 2022. A systematic review : Explainable Artificial Intelligence (XAI) based disease prediction 1, 1–6.
[20] Ali, S., Hafeez, Y., Asghar, S., Nawaz, A., & Saeed, S. (2020). Aspect‐based requirements mining technique to improve prioritisation process: multi‐stakeholder perspective. IET Software, 14(5), 482-492.
[21] Saeed, S. (2023). Digital Workplaces and Information Security Behavior of Business Employees: An Empirical Study of Saudi Arabia. Sustainability, 15(7), 6019.