AN ENSEMBLE LEARNING EXTENSION OF MCAD FOR CYBERATTACK DETECTION IN SDN- BASED HEALTHCARE SYSTEMS

J. Kumari [1], B. Madhu Krishna [2]

Assistant Professor [1], PG Scholar [2]

Department of Master of Computer Applications

QIS College of Engineering & Technology (Autonomous), Ongole

Prakasam Dist., AP

Abstract: Important medical information has to be protected from unwanted access. Healthcare systems make substantial use of SDNs to optimize resource use, security, network administration, and control. SDNs provide many advantages, but because patient data is sensitive, they are susceptible to several threats. These attacks compromise networks and may result in fatal malfunctions. We adapt a layer three (L3) learning switch application to collect normal and abnormal traffic and apply it to the Ryu controller in order to propose a machine learning-based cyber-attack detector (MCAD) for healthcare systems. Our findings enhance the security of healthcare applications and lessen assaults. In order to demonstrate MCAD's advantages and disadvantages, this paper evaluates it using a variety of ML methods and attacks. Excellent reliability for the MCAD is indicated by an exceptional F1-score on both attack and normal classes. The complexity-optimized real-time system of MCAD generates 5,709,692 samples per second.

Index Terms - Network resilience, network management, intrusion detection system (IDS), software defined networking, healthcare, machine learning.

1. INTRODUCTION

Over the past few years, SDNs have been widely utilized across many industries. This is mostly due to the fact that they are a reliable network technology that allows you to manage and administer a network by keeping the control and data planes apart. The SDN architecture provides the controller and applications with more knowledge about the overall status of the network, in contrast to traditional networks that just know about the apps

that are operating on them. Healthcare organizations have begun to employ many of the same off-the-shelf technology, software, and procedures that businesses in other industries use due to the recent fast increase in information and communication technologies (ICT). We anticipated this since medical technology that are networked or connected to the Internet may improve asset management, communications, and electronic health records, all of which save costs. Additionally, the two most crucial considerations for the majority of information systems are the security of systems and devices and the privacy of user data. This is due to the healthcare industry's stringent regulations, which make privacy and safety crucial. Therefore, even if the prices of hospital equipment are predicted, it's crucial that the present McAfee record noted that networked medical instruments may exhibit security flaws as the medical industry attempts to integrate all the technological components of networked infrastructure and operational controls.

By developing a machine learning-based cyber-attack detector (MCAD) that functions in software-defined networks (SDNs), this project aims to improve the security of healthcare systems. A layer three (L3) learning switch program will be used by MCAD, which will be installed on the Ryu controller, to gather and examine both typical and unusual network traffic. A comprehensive picture of the system's performance is provided by the extensive testing conducted in the study using various machine learning algorithms and cyberattack scenarios. MCAD is dependable, as evidenced by its strong F1-score for both normal and attack classes. Additionally, it can process 5,709,692 samples per second in real-time.

One of the biggest issues facing the healthcare industry is protecting sensitive patient data in software-defined networks (SDNs). Although SDNs offer many advantages, they are also susceptible to a variety of threats that might undermine the network and endanger patients. Developing a machine learning-based cyber-attack detector (MCAD) for healthcare systems is the aim of this project. This will be accomplished by using a layer three (L3) learning switch application on the Ryu controller. The goal of this research is to thoroughly assess MCAD's performance against various machine learning algorithms and attack

scenarios in order to enhance network resilience and healthcare data security.

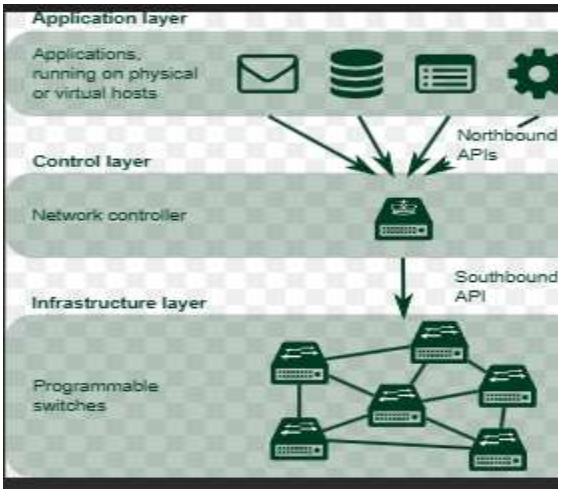

Fig 1 SDN Architecture

2. LITERATURE SURVEY

2.1 Cybersecurity Risks in a Pandemic:

By 2021, cybersecurity risks are predicted to cost the global economy $6 trillion annually, and since COVID-19, the number of assaults has surged fivefold. There is a lot of study on the risks that technology flaws provide to the healthcare sector, but not as much on how hackers might take advantage of pandemics like COVID-19. The reasons why cyberattacks have been especially troublesome during COVID-19 are discussed in this study, along with suggestions for improving patient data protection in the healthcare sector. The Health Insurance Portability and Accountability Act's enforcement has been relaxed by the Office for Civil Rights. While this is helpful when utilizing new platforms like Zoom, it has also weakened technological and physical defenses against hackers. Given that 90% of healthcare providers have previously had data breaches, this is particularly dangerous. Businesses should employ secure networks, such as virtual local area networks, create well-defined software upgrade procedures, and regularly test their systems for vulnerabilities. We can better prepare for the next pandemic by knowing what makes people, healthcare institutions, and companies more vulnerable to cyberattacks.

2.2 A Survey of Ransomware Attacks for Healthcare Systems: Risks, Challenges, Solutions and Opportunity of Research:

One industry that is particularly susceptible to cyberattacks is healthcare. Cybercriminals are attempting to exploit the flaws and security vulnerabilities associated with the industry's rapid growth and transition to digitally enabled healthcare services. The healthcare industry is dealing with a number of extremely potent dangers, including ransomware, as a result of technological advancements. Due to its successful outcomes, ransomware—a

cyberattack that targets businesses and home users—has been more common recently. Over the past few years, the disputes have greatly improved. The study shows an exhaustive survey on Ransomware attacks and fixes these attacks. The main aim of this study is to classify the solution strategies for Ransomware attacks in healthcare that used to prevent the Ransomware, such as Blockchain technology, Software define network technology, Machine Learning, and other tools as well as to highlight many issues faced by researchers during the process of discovering a way to solve Ransomware attacks in health care systems. In addition, the study will provide scientific benefits to researchers in the field of information security, health institutions, and security companies.

2.3 Intelligent Edge Load Migration in SDN-IIoT for Smart Healthcare

The usage of developing technology in the modern period has led to an increase in healthcare problems. Healthcare expenditures can be reduced by combining sensors, the industrial Internet of things (IIoT), and big data analytics to improve patient care. Patients will be able to get more secure, reasonably priced, and advanced medical care as a result. In addition to issues like resource-constrained IoT devices, identity theft attempts, and malevolent insiders, smart healthcare in big data and artificial intelligence requires the use of edge computing services. We are putting out a software-defined networking (SDN)-based security compliance framework for intelligent healthcare load migration systems in order to address these issues. In order to achieve this, experts and researchers are investigating the usage of SDN-IIoT technologies for efficient and instantaneous defense against security threats. Each of the three domains in our suggested design has a single virtual machine and a variety of OpenFlow virtual switches. In order to balance the domain and stop any security threats from occurring during the migration, this scenario aids in moving the healthcare data from the heavily loaded domain to the lighter loaded domain. After collecting the OpenFlow packets via Wireshark, the simulations and efficacy of the performance obtained in the mininet are tested using the RYU SDN controller. The suggested architecture and algorithm provide 80% accuracy for all retrieved healthcare data

packets, enabling secure data management.

2.4 How Secure is the Healthcare Network from Insider Attacks? An Audit Guideline for Vulnerability Analysis:

The availability of wireless connections with modern medical equipment has created several options to give patients improved healthcare support. However, the medical equipment are vulnerable to a number of new assaults due to the flaws in the accessible wireless communication channels. Attacks against medical equipment can be started via a hacked or malware-infected mobile device since smart mobile devices, such computers, tablets, and smartphones, use the same communication channels (WiFi/Bluetooth). A wireless-enabled medical equipment can be used by attackers to obtain private medical information. It is also possible for medical equipment or communication channels to be hacked in order to transmit potentially fatal orders to the devices or provide doctors with inaccurate medical data. Furthermore, it is extremely difficult to stop assaults from compromised mobile devices since they are already inside the security perimeter of a hospital network. In this study, we conduct a methodical analysis of the new threats that might be launched via hacked mobile devices on healthcare networks and equipment. We offer comprehensive audit guidelines to assess a healthcare network's security. We assess a significant university healthcare facility's current security status using our suggested methodology. To lessen some of the potential assaults, we also provide a number of mitigating techniques.

2.5 Cybersecurity Risks and Threats in Healthcare:

Cyber security assaults on healthcare institutions are rising, and the increased use of technology in daily life and healthcare has given hackers additional opportunity to steal and exploit patient data. Cybercriminals can access, utilize, and reveal patient data through EHRs and health information exchange. Cyberattacks on the healthcare sector are rising, thus enterprises must establish a solid cyber defensive system. In this article, we will address current cyber-attack risks on the healthcare system, the present security situation, and ways to improve security. Integrating and improving patient care using EHRs and Health Information Exchange is threatened by

cyber security risks. Security threats have been identified by agencies, notably the US government. Steps have been taken to improve EHR security. They have addressed patient privacy and electronic data storage, but not cyber threats and their potential risks to patient safety and data availability. The rate of technology innovation in healthcare has left the security system behind, and healthcare institutions cannot compete with sophisticated and well-resourced cyber-attack groups. Cyber-attacks are harming patient healthcare in more ways than only privacy, data availability, and integrity. A patient whose medical data were manipulated by a cyber-attack may receive inappropriate treatment or prescriptions, which might impair his health. Healthcare cyber security concerns include ransomware, malware, and deliberate or accidental data leaks. The 2019 Indian clinical research organization attack is an example. Attackers hacked a corporate executive's email account and sent a malware file masquerading as a construction equipment quote. A firm employee opened the email and downloaded the infected file, spreading NetWire malware into the network. A database server with sensitive patient health data was then accessed by intruders. This incident shows the sophistication of cyberattacks and their potential harm to patient treatment.

3. METHODOLOGY

i) Proposed Work:

In order to improve cybersecurity in healthcare systems that depend on Software-Defined Networking (SDN), the suggested solution incorporates MCAD (Machine Learning-based Cyberattack Detector). By implementing an intelligent detection mechanism directly on the Ryu controller, MCAD takes advantage of the centralized control of SDN. It use machine learning algorithms to categorize and react to various cyberthreats in real-time after capturing both legitimate and malicious data via a Layer 3 learning switch. MCAD guarantees prompt threat detection and aids in preserving safe access to sensitive patient data, which is essential in healthcare settings, by examining a variety of attack patterns and traffic trends.

In order to provide high detection accuracy, flexibility, and quick performance, MCAD is trained on a variety of attack datasets and supports a broad range of machine learning models. Because of its architecture,

which enables real-time threat mitigation, the system is appropriate for settings where data integrity and uptime are critical. MCAD exhibits efficiency and dependability with its capacity to handle more than 5 million samples per second and provide a high F1-score in classification. By overcoming the drawbacks of conventional static security solutions like signature-based IDS, this clever, data-driven method enables healthcare SDN systems to identify and react to changing cyber threats dynamically.

ii) System Architecture:

The first step in the MCAD system design is data collection from the network, where topology-based traffic is gathered. To get ready for machine learning, this raw data goes through a number of data preparation procedures, such as cleansing, feature modification, scaling, and shuffling. Following preprocessing, the data is divided into three sets: testing, validation, and training. To create an optimal machine learning model, the training procedure entails model training and hyperparameter tweaking. The Ryu SDN controller then uses this trained model to continually classify incoming unknown network traffic as either regular or attack in real-time. For SDN-based healthcare systems, this architecture guarantees precise, flexible, and effective cyberattack detection.

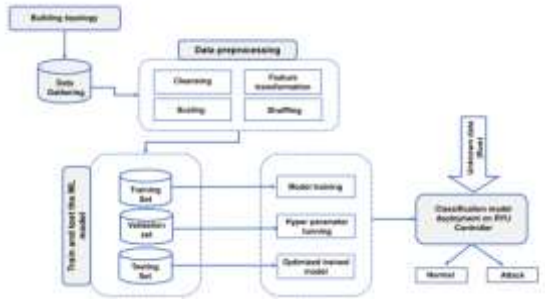

Fig 2 Proposed Architecture

iii) Modules:

a. Proposing a Logical Network Topology: Developing a logical network topology for the healthcare system is the first stage in the model.

b. Data Gathering: In order to train and assess the machine learning (ML) model, the model collects data [19,42]. This covers both standard samples and several types of attacks, including probing attacks, taking use of the VNC port 5900 remote view vulnerability, and taking advantage of the Samba server vulnerability.

c. Data Preprocessing: In order to train the machine learning model, the collected data is cleaned up.

d. Training and Testing the ML Model: To train and assess the ML model, we employ various categorization methods. KNN, decision

tree (DT), random forest (RF), naïve Bayes (NB), logistic regression (LR), adaptive boosting (adaboost), and xgboost (XGB) are some of these. By identifying patterns and minimizing errors, the model generates a mapping function between inputs and outputs. Performance is measured by accuracy [19, 42].

e. Deployment of the project : The user interface makes advantage of the learned machine learning model. This keeps the quality of the healthcare system high by enabling the model to be used in real-time systems.

iv) Algorithms:

K Nearest Neighbour: KNN is a supervised technique that may be applied to both regression and classification. Assuming that similar data points are close to one another in the feature space, it groups data according to the majority class of their k-nearest neighbors (k is specified by the user). In a healthcare SDN setting, KNN may be used to separate various kinds of network traffic [1,8,12]. It assists in identifying anomalous behavior by comparing patterns to known situations.

Decision trees: Decision trees may be used for both regression and classification. They resemble trees, with branches that provide results and nodes that test characteristics. Based on input features, they make decisions by going from the root to the leaves. Decision trees may be used to create rules for identifying oddities on a network. Because decision trees are simple to understand, they are helpful in understanding how the network functions.

Random Forest: A technique called Random Forest creates a single forest by combining many decision trees. You may vote on the trees' estimates or average them to create forecasts. It improves the model's accuracy and lessens overfitting. Random Forest can increase the accuracy of cyberattack detection by aggregating predictions from many decision trees. In healthcare network security, it aids in lowering the quantity of false positives and false egatives [24], [28], and [30].

Naive Bayes: Using Bayes' theorem, Naive Bayes is a kind of probabilistic classifier. It simplifies things by assuming that traits are conditionally independent, which is a popular method for spam filtering and text classification. Text classification, which is essential for identifying poor traffic in healthcare communication,

can be aided by Naive Bayes. It might be used to identify odd textual patterns in network data [54].

Logistic Regression: One kind of statistical model used to address problems with two possible outcomes is called logistic regression. It provides an estimate of the probability that a given input is a member of a particular class. The relationship between the dependent variable (binary outcome) and one or more independent factors is modeled using the logistic function. For binary classification in healthcare network security, logistic regression may be helpful in determining the likelihood that network events are linked to cyberattacks [55].

Adaboost: Adaboost is a technique to create a powerful classifier by combining weak ones. It concentrates on cases that are misclassified, allowing subsequent classifiers to remedy errors. It is frequently used for binary classification. Adaboost is an excellent method for increasing the accuracy of cyberattack detection in healthcare SDNs because it can improve the performance of simple classifiers [56].

XGBoost: XGBoost is a supervised learning technique that employs gradient boosting and is known for being quick, precise, capable of handling missing data, and capable of processing data in parallel. It is widely used in machine learning competitions and applications. You may build a robust and reliable model for identifying cyberattacks using XGBoost, which is known for its exceptional accuracy. This will provide the highest level of security for healthcare data.

Stacking: Stacking uses a meta-learner to generate final predictions based on base classifier outputs. As a result, simple classifiers perform better in terms of prediction. By identifying a greater variety of patterns, it increases accuracy. A collection of several cyberattack detection models that can identify a variety of attack patterns and improve the overall security of healthcare systems may be built via stacking.

Voting: Several simple classifiers' predictions are combined in the voting process. It might be soft (class probabilities) or hard (majority vote). By combining the best features of several models, voting classifiers increase the robustness and accuracy of models. A voting classifier may be used to combine the assessments of

many detection models. This makes it simpler to identify more dependable and powerful cyberattacks in the healthcare network.

4. EXPERIMENTAL RESULTS

To ensure robustness and flexibility, a variety of machine learning methods and cyberattack scenarios were used in the experimental evaluation of the proposed MCAD system. Each model was thoroughly evaluated to determine its efficacy in identifying both typical and malicious traffic after the dataset was split into training, validation, and testing sets. MCAD outperformed the other models in terms of classification ability and reliability, as evidenced by its high F1-score for both attack and normal classes. Additionally, the system's remarkable throughput performance of 5,709,692 samples per second revealed its capacity to function effectively in real-time healthcare settings. These findings demonstrate that MCAD not only offers precise identification but also guarantees low latency and fast processing, both of which are essential for preserving the security and functionality of healthcare networks.

Precision: Precision measures the percentage of samples or incidents that are accurately identified as positives. Therefore, the following formula may be used to determine the precision:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

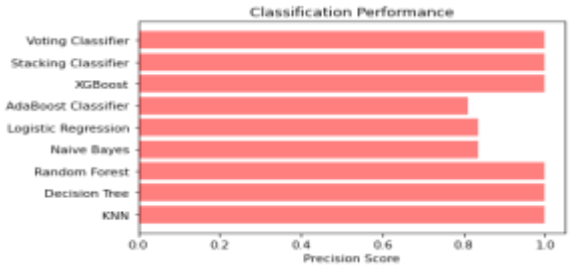

Fig 6 Precision comparison graph

Recall: In machine learning, recall is a statistic that assesses a model's capacity to find all pertinent examples of a given class. It provides information about how well a model captures instances of a certain class and is calculated as the ratio of properly predicted positive observations to all actual positives.

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

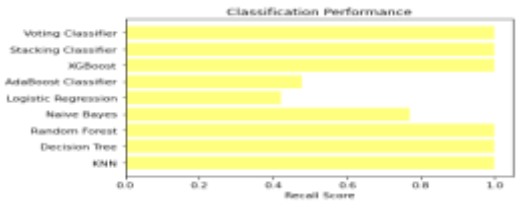

Fig 7: Recall comparison graph

Accuracy: The percentage of accurate predictions in a classification job indicates how accurate a model's predictions are overall.

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

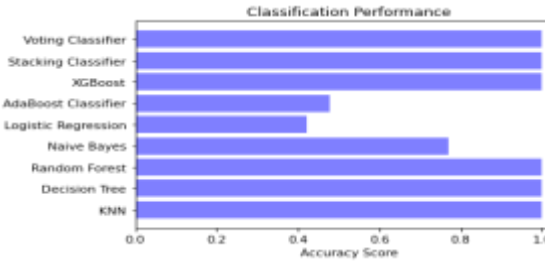

Fig 8: Accuracy graph

F1 Score: The F1 Score is a balanced metric that takes into account both false positives and false negatives, making it appropriate for datasets that are unbalanced. It is calculated as the harmonic mean of accuracy and recall.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

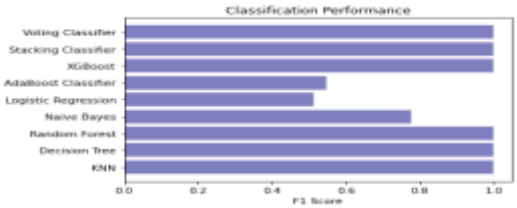

Fig 9: F1Score

ML Model	Accuracy	F1-score	Recall	Precision
KNN	0.999	0.999	0.999	0.999
Decision Tree	0.999	0.999	0.999	0.999
Random Forest	0.999	0.999	0.999	0.999
Naive Bayes	0.770	0.775	0.770	0.834
Logistic Regression	0.421	0.523	0.421	0.834
AdaBoost	0.477	0.548	0.477	0.810
XGBoost	1.000	1.000	1.000	1.000
Stacking Classifier	1.000	1.000	1.000	1.000
Voting Classifier	1.000	0.999	0.999	0.999

Fig 10 Performance Evaluation

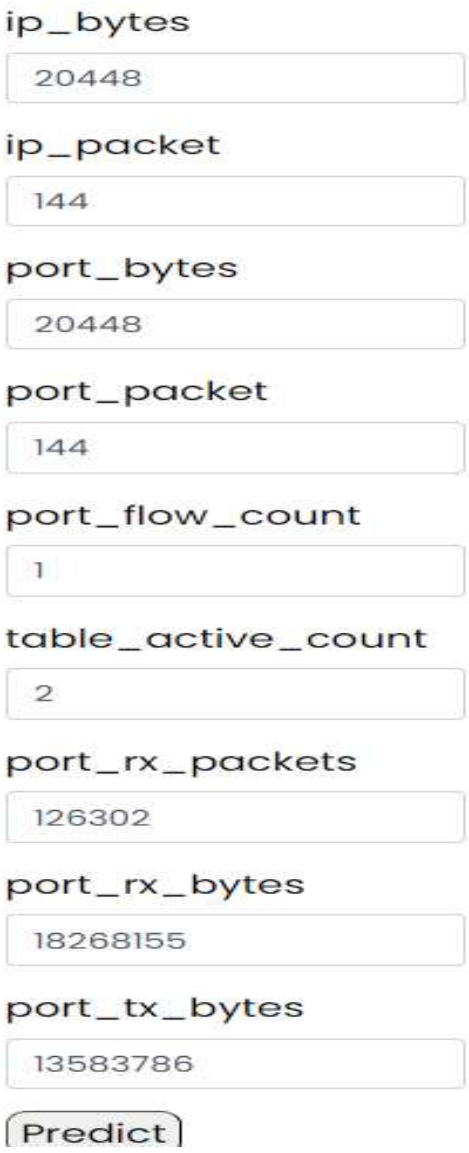

Fig 11: User input

Fig 12: Predict result for given input

5. CONCLUSION

By incorporating machine learning-based threat detection, the suggested MCAD system successfully improves the cybersecurity of healthcare systems using Software-Defined Networking. MCAD quickly and precisely detects

both known and unexpected cyberattacks by utilizing real-time traffic analysis and implementing the optimal model on the Ryu controller. High F1-scores and outstanding throughput attest to the system's robust performance, which validates its dependability and effectiveness in meeting the demanding requirements of healthcare settings. All things considered, MCAD provides a high-performance, scalable, and adaptable solution for safeguarding private medical information and guaranteeing continuous healthcare services.

6. FUTURE SCOPE

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are two deep learning approaches that can be integrated into the MCAD system in the future to increase detection accuracy for complex and dynamic cyber threats. Furthermore, the detection system may be updated in real time without having to retrain from scratch by utilizing adaptive models and online learning. Additionally, the system may be extended to provide cross-domain threat detection, which would make it useful in industries other than healthcare, including banking or industrial IoT networks. Additionally, including automated response systems and blockchain-based tracking for audit trails can greatly improve network resilience and event handling.

REFERENCES

[1] M. Jarschel, T. Zinner, T. Hossfeld, P. Tran-Gia, and W. Kellerer, ‘‘Interfaces, attributes, and use cases: A compass for SDN,’’ IEEE Commun. Mag., vol. 52, no. 6, pp. 210–217, Jun. 2014.

[2] W. Meng, K.-K.-R. Choo, S. Furnell, A. V. Vasilakos, and C. W. Probst, ‘‘Towards Bayesian-based trust management for insider attacks in healthcare software-defined networks,’’ IEEE Trans. Netw. Service Manage., vol. 15, no. 2, pp. 761–773, Jun. 2018.

[3] J. T. Kelly, K. L. Campbell, E. Gong, and P. Scuffham, ‘‘The Internet of Things: Impact and implications for health care delivery,’’ J. Med. Internet Res., vol. 22, p. 11, Nov. 2020.

[4] (2022). Networked Medical Devices: Security and Privacy Threats—Sym antec—[PDF Document]. [Online]. Available: https://fdocuments.
net/document/networked-medical-

devices-security-and-privacy-threatssymantec.html

[5] P. A. Williams and A. J. Woodward, ‘‘Cybersecurity vulnerabilities in medical devices: A complex environment and multifaceted problem,’’ Med. Devices, Evidence Res., vol. 8, pp. 305–316, Jul. 2015.

[6] C. M. Williams, R. Chaturvedi, and K. Chakravarthy, ‘‘Cybersecurity risks in a pandemic,’’ J. Med. Internet Res., vol. 22, no. 9, Sep. 2020, Art. no. e23692.

[7] N. Thamer and R. Alubady, ‘‘A survey of ransomware attacks for healthcare systems: Risks, challenges, solutions and opportunity of research,’’ in Proc. 1st Babylon Int. Conf. Inf. Technol. Sci. (BICITS), I. Babil, Ed., Apr. 2021, pp. 210–216.

[8] H. Babbar, S. Rani, and S. A. AlQahtani, ‘‘Intelligent edge load migration in SDN-IIoT for smart healthcare,’’ IEEE Trans. Ind. Informat., vol. 18, no. 11, pp. 8058–8064, Nov. 2022.

[9] R. Hasan, S. Zawoad, S. Noor, M. M. Haque, and D. Burke, ‘‘How secure is the healthcare network from insider attacks? An audit guideline for vulnerability analysis,’’ in Proc. IEEE 40th Annu. Comput. Softw. Appl. Conf. (COMPSAC), Jun. 2016, pp. 417–422.

[10] (Apr. 2015). 92% of Healthcare IT Admins Fear Insider Threats Thales. Accessed: Mar. 21, 2023. [Online]. Available: https://cpl.thalesgroup.com/about-us/newsroom/news-releases/92-healthcare-it-admins-fearinsider-threats

[11] D. Chaulagain, K. Pudashine, R. Paudyal, S. Mishra, and S. Shakya, ‘‘OpenFlow-based dynamic traffic distribution in software-defined networks,’’ in Mobile Computing and Sustainable Informatics. Singapore: Springer, Jul. 2021, pp. 259–272.

[12] R. Khondoker, A. Zaalouk, R. Marx, and K. Bayarou, ‘‘Feature-based comparison and selection of software defined networking (SDN) controllers,’’ in Proc. World Congr. Comput. Appl. Inf. Syst. (WCCAIS), Jan. 2014, pp. 1–7.

[13] Z. Ghaffar, A. Alshahrani, M. Fayaz, A. M. Alghamdi, and J. Gwak, ‘‘A topical review on machine learning, software defined networking, Internet of Things applications: Research limitations and challenges,’’ Electronics, vol. 10, no. 8, p. 880, Apr. 2021, doi: 10.3390/electronics10080880.

[14] C.-S. Li and W. Liao, ‘‘Software defined networks [guest editorial],’’ IEEE Commun. Mag., vol. 51, no. 2, p. 113, Feb. 2013.

[15] M. H. Rehmani, A. Davy, B. Jennings, and C. Assi, ‘‘Software defined networks-based smart grid communication: A comprehensive survey,’’ IEEE Commun. Surveys Tuts., vol. 21, no. 3, pp. 2637–2670, 3rd Quart., 2019.

Author Profiles

Mrs. Jasti kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Mr. B. MADHU KRISHNA is an MCA Student in the Department of Computer Application at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He has Completed Degree in B.Sc.(computers) from Sai Madhavi Degree collage Anaparti, East Godavari district.