IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

Intelligent Detection System for Multi-Step Cyber Attack Based on Machine Learning

Sunil J, Darshan S B, Darshan M R

Assistant Professor, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

ABSTRACT: Cyberattacks use malware to carefully alter computer systems and networks in order to stifle processes and activities, converge data, or limit data access. These kinds of attacks have significantly increased over time. Advanced defensive techniques are required because of the increase in complexity and structure. Traditional techniques for detecting cyberattacks are no longer viable in the face of increasing security threats. This study proposes an intelligent intrusion detection system. Additionally, the proposed system aims to assess the k-nearest neighbour algorithm's (KNN) ability to differentiate between tampered and real data. The Multi-Step Cyber-Attack Dataset (MSCAD), a trustworthy dataset, is used to ascertain the behaviour of the novel attack types. Additionally, the model was trained using 60% of the dataset, with the remaining 40% being used for testing. Evaluation criteria such as F1 score, recall, accuracy, and precision are employed. The suggested system-based KNN may improve detection performance, according to experiments. Furthermore, the recommended method reduces false alarms while improving detection accuracy.

KEYWORDS: Cyber-attacks, intrusion detection system, KNN, MSCAD, accuracy.

I. INTRODUCTION

In a cyberattack, malware is purposefully used to alter data or impede operations on computer systems and networks. Attacks of this kind have significantly increased over time. Because of the increased structure and complexity, a greater level of protective techniques and innovative detection is needed. Traditional methods of detecting cyberattacks are ineffective as security threats increase.

[1] Cybersecurity has grown to be a major problem in this technological age. Due to the widespread use of networking devices, intrusion defence has grown in significance.

[2] Hackers' destructive attacks are identified by the Cyber Attack Detection Model (CADM). Similar to a CADM, an intrusion detection system (IDS) examines network traffic data for unusual content. The widespread use of wireless devices and the horrifying atrocities perpetrated by attackers have made CAM a crucial component of network security design.[3,4] The Internet of Things (IoT) has changed people's lives. IoT networks impact well-known companies and industries in addition to implementing eHealth, intelligent transportation, and smart environments (homes and cities). The Internet of Things presents growing security vulnerabilities in spite of its remarkable scalability and agility.

[5] Cybersecurity is crucial for emerging technologies including information processing, communications networks, and Internet of Things networks. In addition, barcodes, cloud computing, and social networks are important IoT technologies that present cyber-security issues. Since smart homes are among the most susceptible to cybercrime, many nations have researched IoT applications. To safeguard IoT networks, they have given top priority to putting certain laws and guidelines into place. Cybersecurity concerns rise as an IoT network's linked devices grow in number.

[6] Presenting an intelligent intrusion detection system for identifying novel attack types is the paper's contribution. Additionally, the suggested approach made use of a trustworthy dataset that had been trained with a KNN classifier. The following are the remaining sections of this paper: Section III presents some information about cyberattacks, whereas Part II shows the related work. The suggested system's technique is presented in Section IV. Section V discusses the findings and debate. Lastly, section VI emphasizes the conclusion.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

II. LITERATURE SURVEY

Due to the rapid growth in the number of apps and networks, cyber multi-step attacks have increased. Consequently, there is a growing need for a trustworthy solution. Limitations and difficulties encountered, like errors in prediction and issues with large amounts of data, were briefly discussed in certain studies. A fully learning-based assault detection device version for energy structures was created in [5,6,7] using logs and data obtained through phasor size units to train the device.

The AWV model was able to identify 37 distinct types of power grid activities with accuracy because to the data processing technique, which also made the model more exact. The data was subjected to a number of machine learning models, with AdaBoost's basic classifier being the random forest. The suggested model was compared to others using a number of comparison criteria. This model has an accuracy rate of 93.91%, an identification rate of 93.6% higher than 8%, and a rate of 93.36% higher than 8%, based on the testing findings.

The issue of storing large and varied datasets has been addressed by cloud computing in [8]. A distributed IDS technique is used to manage massive volumes of warning data. Spark was necessary for the investigation. Additionally, they trained the model using naive Bayes classification.

Because of large data, spoofing, DDoS, and MiTM attacks have increased in frequency [9]. Data dimensionality was decreased using the data dimensionality reduction approach, which increased the detection rate. Both standalone classifiers, SVM and NNet, as well as ensemble techniques, XGBoost and CTree, were used. The authors in [10] used the Chi-Square test to determine the key properties for their intrusion detection algorithm. The ensemble approach made use of a supervised classification method (SVM, modified naive Bayes, and LPBoost).

Lightweight intrusion detection systems are necessary for real-world applications in [11]. The authors employed a wrapper attribute selection method to extract the attributes. To differentiate between legitimate network traffic and malicious traffic, apriori association rule mining and support vector machines are employed. In order to distinguish between network attacks and equipment failure, the impact on communication was examined in [12].

By connecting diverse abnormality sources, a machine learning-based architecture was created to solve the differentiation challenge. It has proven to have a favorable classification performance of more than 85% based on simulation and real-time testing.Based on functionality and experimental findings, a quantitative description of the studied classes was obtained, and a method for improving classification accuracy was proposed.

In this work, we offer an intelligent intrusion detection system that uses the MSCAD dataset to adopt the KNN classifier and determine attack behavior. The findings demonstrate that the recommended system improved the detection accuracy rate.

III. CYBER ATTACKS

An attack launched from at least one computer against another computer, many computers, or various IT infrastructures is known as a cyber-attack. These attacks fall into two broad categories: attempts to damage the target computer with the intention of gaining access to its data and potentially obtaining administrator privileges on it [13]. Cyberattacks can take many different forms. Identifying decoded movements of sensitive and crucial information is one of the core components of dynamic organization assaults. However, pursuing inactive targets sometimes involves keeping an eye on links with dangerous organizations. These assaults can target any Internet of Things device. They are used to cause harm, weaken system control, or obtain sufficient access to the victim's personal information [14,15]. On a larger scale, organizations are targeted to prevent residents from receiving power or water. Homegrown models recall attacks on house computerization systems, which attackers use to control lighting, heating, cooling, and security systems. Without identifying the organization's weaknesses, it is difficult to anticipate a possible attack in the field of network security. Accordingly, it is essential to identify and comprehend the various current digital attack tactics in order to strengthen organizational vulnerabilities.

Cyber attacks come in many forms, such as malware, zero-day attacks, denial-of-service attacks, phishing, session hijacking, ransomware, SQL injection, man-in-the-middle attacks, and password attacks. One of the most worrying

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

issues is security; any cyberattack, no matter how big or little, begins with the deception of a weak link in a security system, and programmers are adept at identifying these weak areas and using them to their advantage. Anything connected to an organization is likely to be targeted online. They can accomplish this more easily than at any other time in recent memory in the rapidly developing Internet of Things. Because IoT devices are individually connected, each programmer must identify one vulnerability and handle the entirety of the data. Knowing the network's vulnerabilities is important for detecting or monitoring cyberattacks. In order to identify the concept of the information at risk and comprehend the cause of the attack, it is also critical for the digital security partners to learn the attacker's motivations.

IV. METHODOLOGY

In order to guarantee security against cyberattacks in smart settings, a detection model is employed in this paper. The suggested approach for intrusion detection is based on the KNN classifier, as seen in Figure 1. Following dataset selection, the pre-processing phase is used to convert some symbols to numbers and boost the effectiveness of the detection rate. The training phase follows, during which 60% of the dataset is used to train the model using the KNN. The testing phase is the last stage, during which the model is evaluated using the 40% of data that is still available. In order to gauge the effectiveness of the suggested system, the accuracy detection rate and the four different kinds of alarms are also computed at this phase.

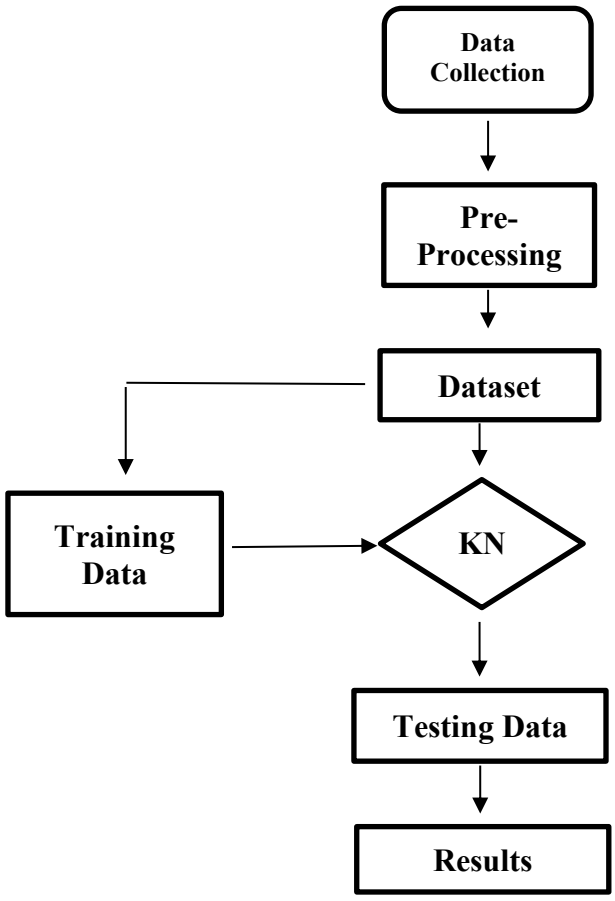

Fig 1. The flowchart of the proposed system

Since the features are composed of characters, symbols, and numbers, the preprocessing stage is crucial after dataset collection. All that KNN deals with throughout the training and testing stages is numerical data. In order to convert these characters and symbols to numeric values, the preprocessing step is used. When creating an intelligent model, the training and testing stage is crucial. KNN was used in a training phase to increase accuracy, decrease, and rate the rate of false alarms. The training set, which makes up 60% of the total, and the testing set, which makes up 40% of the total,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

are the two subsets of the dataset. The testing phase is thought to be necessary in order to assess the accuracy of the model. Metrics including precision, recall, and F1-score are calculated during this stage.

A. Dataset Description

Establishing a strong intrusion detection system requires a current and trustworthy dataset to ascertain how different kinds of attacks behave. The Multi-Step Cyberattack Dataset (MSCAD) is used in this paper [16]. This dataset consists of seven files:

N-0, Scan-1, MSCAD.xlsx, App-01, App-02, W-B-01, W-B-02

MSCAD.xlsx contains the labelled version of the dataset. Wireshark was used to process the six PCAP files. The timestamp of the network communication is assessed in order to classify malicious and normal network traffic. Following processing of these PCAP files, 77 features with labels were included in the resulting MSCAD dataset. Password cracking is the initial attack scenario in MSCAD, while Distributed Denial of Service (DDoS) is the second attack type.

B. K-Nearest Neighbours (KNN)

KNNs can be applied directly to both regression and classification issues. It is simple to comprehend and put into practice. However, a big challenge is that it becomes noticeably slower as the amount of data being used grows [17]. The KNN classifier does not need a

Since the data itself is a model that will be utilized for future detection, the training phase is extremely timeefficient wh en it comes to improvising for random modeling on the provided data.[18]

C. EVALUATION METRICS

Evaluation metrics are used by an intelligent intrusion detection system to assess its classifiers. Metrics are used to assess the effectiveness of an intrusion detection system [19]. Accuracy has been used in this work to analyze the effectiveness of using KNN. The ratio of successfully classified activities to all recognized activities can be used to define accuracy. Equation 1 is used to calculate it:

Accuracy = correctly classified activities / total number of activities (1)

To measure and evaluate system performance, four types of alarms (confusion matrix) should be calculated, A true positive is a true positive, a false positive is a false positive, a true negative is a true negative, and a false negative is a false negative [20]:

TP=TP/TP+FN (2)

TN=TN/TN+TP (3)

FN=FN/FN+TP (4)

FP=FP/FP+TN (5)

Furthermore, precision and recall are computed using Equations (6 - 7).

precision=TP/TP+FN (6)

recall=TP/TP+FP (7)

Another utilized metric is the of F1-score which is considered a harmonic average of recall , and a precision. F1-score is computed following eq. 8:

F1- score =2* (precision* recall/ precision+ recall) (8)

V. EXPERIMENTAL RESULTS

This work separates the dataset into training and testing sets using 60% and 40% ratios. The suggested intrusion detection system's accuracy, precision, recall, and F1 score are assessed. Our tests are carried out on a 2x 2.4GHz Intel Xeon 8 Core computer with two Gigabyte Dell PowerEdge T430 GPUs (RAM) and 32 GB of DDR4 memory.

Table 1 displays the findings for both the confusion matrix rate and the detection accuracy rate.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

TABLE 1 THE CONFUSSION MATRIX AND THE ACCURACY VALUE

Alarms Type	Value
TP	**94.21%**
TN	**90.86%**
FP	**9.14%**
FN	**5.79%**
Accuracy	**85.59%**

Table 1 shows that the training's accuracy rate is 82.59%. This suggests that the KNN machine-learning technique has produced findings that are widely recognized for identifying various attack types. The analysis of the used datasets demonstrates that the MSCAD dataset enables us to evaluate the suggested system's capacity to identify instances of intrusion. Our findings are contrasted with those of earlier studies [21, 22].

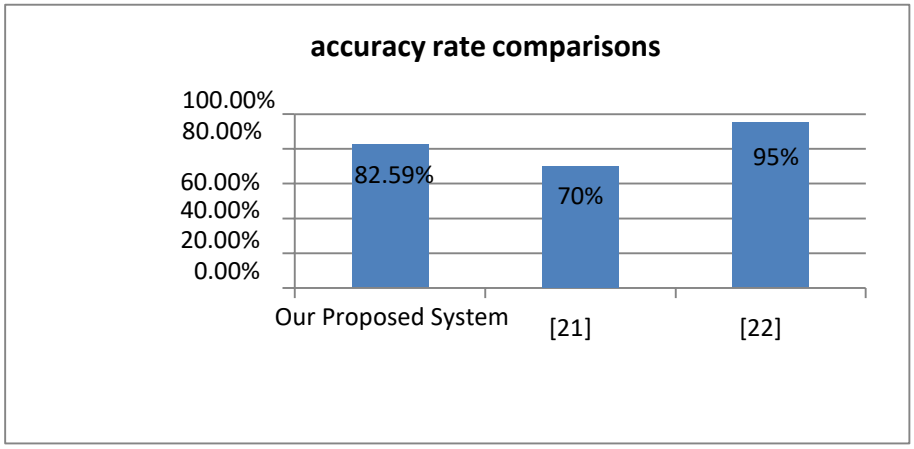

Fig 2. Accuracy rate comparisons

Figure 2's compression demonstrated our suggested system's accepted accuracy rate.

TABLE 2. EVALUATION METRICS

Evaluation metrics	Value
Precision	94.21%
Recall	91.1%
F1-score	92.6%

The performance indicators listed in Table 2 demonstrate the extent to which the suggested system can identify novel attack types.

VI. CONCLUSION

This study looked into the ML-based intrusion detection method for cyber security threats. KNN is the machine learning approach used in this investigation. The suggested system was assessed using the confusion metrics (TP, TN, FP, and FN). Accuracy, precision, recall, and F1-score were computed. Acceptable results were conveyed by the investigation's findings.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311033|

REFERENCES

[1]. Mohammed, A. H. K., Jebamikyous, H. H., Nawara, D., & Kashef, R. (2021). IoT cyber-attack detection: A comparative analysis. *ACM International Conference Proceedings Series*, 117–123. doi: 10.1145/3460620.3460742.
[2]. Alqahtani, H., Sarker, I. H., Kalim, A., Hossain, S. M. M., Ikhlaq, S., & Hossain, S. (2020). Cyber intrusion detection using machine learning classification techniques. *Springer Singapore*, vol. 1235 CCIS, July.
[3]. Palop, J. J., Mucke, L., & Roberson, E. D. (2010). Chapter 17 and Neuronal Network Hyperexcitability in Mouse Models. *In Mouse Models of Cognitive Dysfunction*, vol. 670, 245–262. doi: 10.1007/978-1-60761-744-0.
[4]. Ahmad, B., Jian, W., & Ali, Z. A. (2018). Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions. *Journal of Computer Networks and Communications*, 2018. doi: 10.1155/2018/6383145.
[5]. Primartha, R., & Tama, B. A. (2018). Anomaly detection using random forest: A performance revisited. *Proceedings of the 2017 International Conference on Data and Software Engineering (ICoDSE)*, 2018-Janua, 1–6. doi: 10.1109/ICODSE.2017.8285847.
[6]. 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)*, vol. 7, p. 100059. doi: 10.1016/j.iot.2019.100059.
[7]. Bouhlel, M. S., & Rovetta, S. (2018). Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, vol. 1.
[8]. Bansal, A., & Kaur, S. (2019). Data dimensionality reduction (DDR) scheme for intrusion detection system using ensemble and standalone classifiers. *Springer Singapore*, vol. 1045.
[9]. Thaseen, I. S., Kumar, C. A., & Ahmad, A. (2019). Integrated Intrusion Detection Model Using Chi-Square Feature Selection and Ensemble of Classifiers. *Arabian Journal for Science and Engineering*, vol. 44, no. 4, 3357–3368. doi: 10.1007/s13369-018-3507-5.
[10]. Sarumi, O. A., Adetunmbi, A. O., & Adetoye, F. A. (2020). Discovering computer networks intrusion using data analytics and machine intelligence. *Scientific African*, vol. 9, p. e00500. doi: 10.1016/j.sciaf.2020.e00500.
[11]. Tertytchny, G., Nicolaou, N., & Michael, M. K. (2020). Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine learning. *Microprocessors and Microsystems*, vol. 77, p. 103121. doi: 10.1016/j.micpro.2020.103121.
[12]. Mohamed, A., & Køien, G. (2015). Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders, and Attacks. *Journal of Cyber Security*, vol. 4, no. 10, 65-88.
[13]. Cheng, Y., Naslund, M., Selander, G., & Fogelström, E. (2012). Privacy in machine-to-machine communications: A state-of-the-art survey. *International Conference on Communication Systems (ICCS)*, 75–79.
[14]. Rudner, M. (2013). Cyber-threats to critical national infrastructure: An intelligence challenge. *International Journal of Intelligence and CounterIntelligence*, vol. 26, no. 3, 453–481.
[15]. Almseidin, M., Al-Sawwa, J., & Alkasassbeh, M. (2022). Generating a Benchmark Cyber Multi-Step Attacks Dataset for Intrusion Detection. *2022*, 1–15.
[16]. Govindarajan, M., & Chandrasekaran, R. (2009). Intrusion detection using k-Nearest Neighbor. *2009 First International Conference on Advanced Computing*, 13–20. doi: 10.1109/ICADVC.2009.5377998.
[17]. Liao, Y., & Vemuri, V. R. (2002). Use of K-Nearest Neighbor classifier for intrusion detection. *Computers & Security*, vol. 21, issue 5, 439–448.
[18]. Wang, W., Xu, H., Alazab, M., Gadekallu, T. R., Han, Z., & Su, C. (2021). Blockchain-based reliable and efficient certificateless signature for IIoT devices. *IEEE Transactions on Industrial Informatics*.
[19]. Wang, W., Xu, H., Alazab, M., Gadekallu, T. R., Han, Z., & Su, C. (2021). Blockchain-based reliable and efficient certificateless signature for IIoT devices. *IEEE Transactions on Industrial Informatics*.
[20]. Alheeti, K. M. A., Gruebler, A., & McDonald-Maier, K. D. (2015). On the detection of grey hole and rushing attacks in self-driving vehicular networks. *2015 7th Computer Science and Electronic Engineering Conference (CEEC)*, 231–236.
[21]. Ofori, A. Y., Swart, C., Boateng, F. A. O., & Islam, S. (2022). Cyber resilience in supply chain system security using machine learning for threat prediction. *Continuous Resilience Review*, vol. 4, 1–36.
[22]. Rashid, M. M., Kamruzzaman, J., Hassan, M. M., Imam, T., & Gordon, S. (2020). Cyberattacks detection in IoT-based smart city applications using machine learning techniques. *International Journal of Environmental Research and Public Health*, vol. 17, 9347.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details