IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 497-502

DOI: 10.17148/IARJSET.2023.10774

Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions

Prajwal Gowda S[1], C S Swetha[2]

Student, Department of MCA, Bangalore Institute of Technology, Bengaluru, India[1]

Assistant Professor, Department of MCA, Bangalore Institute of Technology, Bengaluru, India[2]

Abstract: For Agriculture 4.0 cyber security, the system, evaluate, and analyse intrusion detection systems. We discuss cyber security risks as well as assessment criteria that were employed in the performance evaluation of an intrusion detection system for Agriculture 4.0. Then, we assess intrusion detection systems in light of upcoming technologies such as cloud computing. Intrusion detection is a critical security issue in today's cyber environment. A large range of strategies based on machine learning methodologies have been developed. So, in order to detect the infiltration, we created machine learning algorithms. A network-based intrusion detection system (NIDS) is often installed at network points such as gateways and routers to detect network traffic intrusions. We deliver a complete report based on the machine learning approach employed. We emphasise the problemsand future research areas for Agriculture 4.0 cyber security intrusion detection. IDSs employ artificial intelligence-based approaches such as machine learning and cloud computing to identify harmful conduct. Finally, we can identify the IDS using machine learning and save the observed data in free cloudstorage.

I. INTRODUCTION

Agriculture 1.0, Agriculture 2.0, Agriculture 3.0, and Agriculture 4.0 are the four generations of the agricultural and industrial revolution. Agriculture 1.0 refers to agricultural practises from the dawn of human civilization until the end of the nineteenth century, when farmers relied heavily on traditional cultivation tools like the traditional plough to create favourable conditions for seed placement and plant growth. Agriculture 2.0 was the name given to a growth in agricultural productivity at the beginning of the twentieth century that was based on agricultural technology such as combines, irrigation, harvesting, trucks, tractors, aeroplanes, helicopters, and so on. Agriculture 3.0 emerged in the early 1970s and is based on green renewable energy such as bio energy, geothermal energy, solar energy, and hydropower. The term "Agriculture 4.0" emerged after "Industry 4.0," which is defined by a combination of emerging technologies such as block chain, software defined networking (SDN), artificial intelligence, Internet of Things.

(IoT), IoT devices, 5G communications, drones, fog/edge computing, cloud computing, network function virtualization (NFV), smart grids, and so on. These developing technologies have been widely employed in Industry 4.0, and their implementation in agricultural contexts is not difficult to duplicate. As a result, the primary difficulty of establishing Agriculture 4.0 is not the deployment of emerging technologies, but rather the assurance of security and privacy, given that the deployment of thousands of IoT-based devices is in an open field. Furthermore, there are several security and privacy concerns connected with each tier of the system. An opponent, for example, can launch several cyber-attacks, such as distributed denial-of-service (DDoS) assaults, to disrupt a service and subsequently insert fake data, affecting food safety, agri-food supply chain efficiency, and agricultural production. The cyber security research community recommends the usage of intrusion detection systems (IDS), a network security technology dedicated to continually watching events inside a computer or networking system and comparing them to intrusion evidence. IDSs utilise artificial intelligence-based techniques to identify malicious behaviour, such as hybrid machine learning, voting. based extreme learning machines, deep learning techniques, hierarchical approaches, reinforcement learning, and so on. Many surveys have addressed IDSs based on machine learning techniques.

Objectives: The primary goal of our study is to properly identify, forecast, and detect IDS.To improve speed, implement classification methods.

- To save the identified data in a free cloud storage service.
- To improve classification algorithms' overall performance.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 497-502

DOI: 10.17148/IARJSET.2023.10774

II. LITERATURE SURVEY

2017 study on SDN-based network intrusion detection systems employing artificial intelligence techniques Methodology: [1] Software Defined Networking (SDN) offers the opportunity to detect and monitor network security issues caused by the advent of programmable features. To secure computer networks and address network security concerns, Machine Learning (ML) methods have recently been incorporated in SDN-based Network Intrusion Detection Systems (NIDS). In the context of SDN, a stream of advanced machine learning methodologies- deep learning technology (DL) - is beginning to develop.

We evaluated different current research on machine learning (ML) methodologies that use SDN to create NIDS in this study. More particular, we assessed deep learning strategies for constructing SDN-based NIDS. Meanwhile, in this survey, we discussed techniques for developing NIDS models in an SDN context. This survey concludes with a discussion of existing issues and future work in implementing NIDS using ML/DL.

- Statistical approaches do not need prior knowledge of network assaults.

- The primary drawbacks of many feature learning methods are their complexity and high implementation costs.

[2] A rigorous survey on multi-step attack detection was conducted in 2018. Cyber-attacks have posed a hazard to individuals and companies since the inception of the Internet. They have grown in complexity with computer networks. In order to achieve their ultimate goal, attackers must now go through many intrusive procedures. The collection of these processes is referred to as a multi-step assault, multi-stage attack, or attack scenario. Becausethe correlation of more than one activity is required to comprehend the assault plan and identify the danger, their multi-step nature makes the danger, their multi-step nature makes intrusion detection difficult. Since the early 2000s, the security research community has attempted to provide ways to identify this type of danger and forecast further moves. The goal of this study is to collect all articles that provide multi-step assault detection systems. We concentrate on approaches that Bibliographic research is used to find relevant publications. Our efforts result in a corpus of 181 papers describing and categorising 119 approaches. The publication analysis allows us to draw some conclusions about the level of research in multi-step assault detection

- The benefit of this system is that it detects harmful network events using IDS signatures and tracks their progression as successive events, looking for matches in terms of IP address or port

- Because an attacker does not need to follow a certain order when performing a multi-step assault, the collection of alternative action sequences might be extremely complicated. [3] Clustering-based real-time anomaly detection—a big data technology breakthrough in 2019 Lately, the ever-increasing use of linked Internet-of-Thingsdevices has increased the volume of real-time network data at a rapid pace. At the same time, network attacks are unavoidable; hence, detecting abnormalities in real-time network data has become critical. K-means, hierarchical density-based spatial clustering of applications with noise (HDBSCAN), isolation forest, spectral clustering, and agglomerative clustering are used to undertake critical comparative analysis. When compared to other algorithms, the evaluation results demonstrated the usefulness of the suggested framework with a considerably better accuracy rate of 96.51%.. Furthermore, the proposed framework outperforms current algorithms in terms of memory use andexecution time. Finally, the suggested approach allows analysts to track and spot abnormalities in real time. The spark iterative computation architectural allows large-scale machine learning algorithms to reach high levels of efficiency in results, and the spark.ml API for pipeline provides developers with a diverse set of new modules to interact with their architecture.

- Small and slow-ramped attacks can avoid statistical tactics by limiting the impact of the attack below statistical criteria. [4] Machine Learning Techniques for Network Intrusion Detection Evaluation, 2018 A network traffic abnormality may suggest a probable network breach, hence anomaly detection is critical for detecting and preventing security assaults. The majority of the early research in this field. and commercially available Intrusion Detection Systems (IDS) are signature-based. The issue with signature-based methods is that the database signature must be updated when new attack signatures become available, making themunsuitable for real-time network anomaly detection. Machine learning classification approaches have recently become popular in anomaly detection. We apply and analyse seven alternative machine learning approaches with information entropy computation to the Kyoto 2006+ data set. Our data indicate that, for Advantage:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 497-502

DOI: 10.17148/IARJSET.2023.10774

- Training time is limited.

Disadvantage: The fundamental disadvantage of the signature-based technique is that the database signature must be updated when new signatures become available, making it unsuitable for real-time network anomaly detection. Comparing the outcomes of the seven methods mentioned here using numerous performance criteria is quite tricky.

[5] Exploring the Shodan via the Lens of Industrial Control Systems The Industry Control System (ICS), as a crucial component of critical infrastructure, is increasingly vulnerable to cyber assaults. The appearance of the Shodan search engine heightened the threat. The Shodan search engine has become a favourite toolbox for attackers and penetration testers due to its ability to find and index Internet-connected industrial control equipment.In this work, we employ honeypot technology to undertake a thorough investigation of the Shodan search engine. We begin by deploying six distributed honeypot systems and collecting three months' worth of traffic data. To investigate Shodan, we create a hierarchical DFA-SVM identification model to identify Shodan scans based on function code and traffic characteristic, which is then customised to locate Shodan and Shodan-like scanners that are superior to the original. , we undertake a thorough study of Shodan scans and assess the influence of Shodan onindustrial control systems in terms of scanning duration, frequency, scanning port, area preferences, ICS protocol preferences, and ICS protocol function code proportion. As a result, we present several protective methods to lessen the Shodan danger.

The key benefit of SVM is that it is a machine learning model with a high detection rate of tiny samples and a good generalisation ability, making it suited for handling high-dimensional and non-linear Shodan traffic from a limited number of Shodan scanners. Disadvantage: • Prediction is not precise.

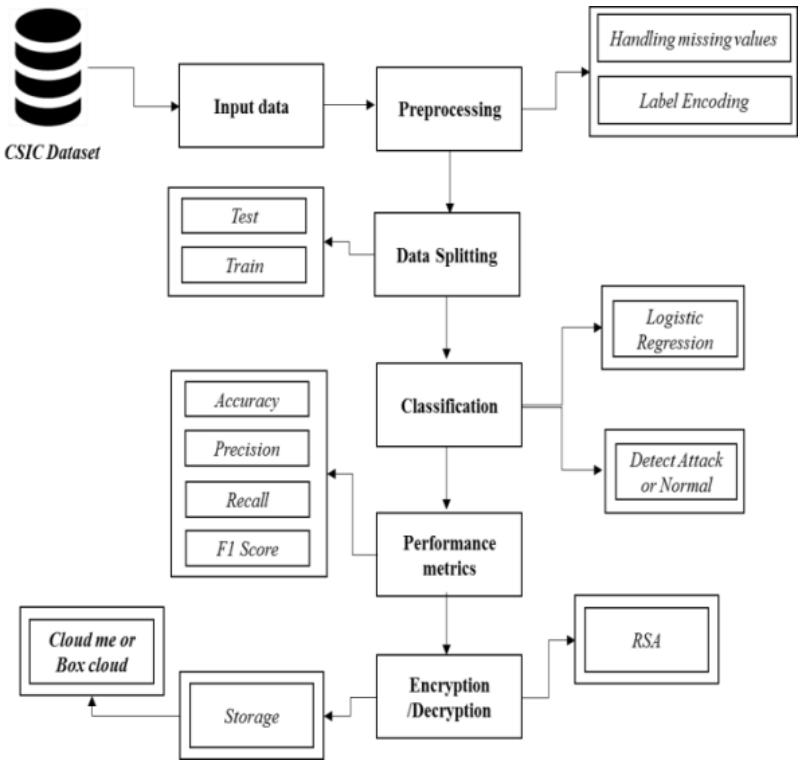

Fig [1] System Architecture

III. EXISTING SYSTEM

We study and analyse intrusion detection technologies for agricultural cyber security in current systems. We discuss cyber security risks as well as assessment criteria that were employed in the performance evaluation of an intrusion detection system for Agriculture 4.0. Then, we assess intrusion detection systems in light of upcoming technologies such as cloud computing, fog/edge computing, and network virtualization. We present a detailed classification of intrusion detection systems in each developing technology based on the machine learning approach utilised.

ISO 3297:2007 Certified

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 497-502

DOI: 10.17148/IARJSET.2023.10774

In addition, we discuss publicly available datasets as well as the implementation frameworks used in the performance evaluation of intrusion detection systems for Agriculture 4.0. Finally, we discuss the obstacles and future research objectives for Agriculture 4.0 cyber security intrusion detection. DISADVANTAGES: • It is inefficient for huge volumes of data; • It has theoretical limits.

- Training duration is extensive.
- The procedure is carried out without the removal of unnecessary data.

IV. PROPOSED SYSTEM

The CSIC_2020dataset was used as input in this system. The dataset repository was used to obtain the input data. Then we must carry out the data pre-processing stage. In this stage, we must manage missing values to avoid incorrect prediction and encode the label for input data.The dataset must then be divided into two parts: test and train. The data is being separated depending on a ratio. The majority of the data will be present in train. A reduced fraction of the data will be present in the exam. The training phase is used to assess the model, whereas the testing phase is used to forecast the model. The classification algorithm (i.e., machine learning algorithm) such as Logistic regression must next be implemented, followed by encryption techniques such as RSA. Finally, the experimental findings suggest that various performance measurements, such as accuracy, may be stored in the cloud for free.

ADVANTAGES:

- It is efficient for a big number of datasets;
- It consumes little time.
- The procedure begins with the removal of undesirable data.

V. IMPLEMENTATION

MODULES:

Data selection Pre-processingData splitting Classification Prediction

Data selection

The input data were obtained from a dataset source.

- The CSIC_2020 dataset is used in our method.
- Data selection is the process of anticipating an IDS assault.
- The input dataset was obtained from a source such as the UCI repository.
- The dataset contains data such as the URL, method, categorization, and so on.
- We can read or load our input dataset in Python using the panda module.
- Our dataset is in the '.csv' format.

Pre-processing

Data pre-processing is the process of deleting undesirable data from a dataset. • Pre-processing data transformation techniques are used to turn the dataset into a machine learning-friendly structure**.** • This process also includes cleaning the dataset by deleting extraneous or damaged data that might impair the dataset's correctness, making it more efficient.

ISO 3297:2007 Certified

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 497-502

DOI: 10.17148/IARJSET.2023.10774

- Elimination of missing data
- Categorical data encoding
- Missing data removal: This method replaces null values such as missing values and Nan values with 0.
- Any missing or duplicate values were eliminated, and the data was cleansed of any irregularities.
- Categorical data encoding: Variables having a finite number of label values are used to represent categorical data.
- The fact that the vast majority of machine learning algorithms require numerical input and output variables.

Data splitting

- Data are required during the machine learning process in order for learning to occur. In addition to the data necessary for training, test data are required to assess the algorithm's performance and determine how effectively it performs.
- We regarded 60% of the input dataset to be training data and 40% to be testing data in our procedure. Data splitting is the process of dividing accessible data into two halves, typically for cross-validation reasons. One portion of the data is used to create a predictive model, while the other is utilised to assess the model's performance. Part of analysing data mining models is separating data into training and testing sets. Normally, when you divide a data collection into

Classification

We can use machine learning algorithms like Logistic Regression in our approach. Logistic regression is a statistical analytic approach that uses past observations of a data set to predict a binary result, such as yes or no. A logistic regression model forecasts a dependent variable by examining the connection between one or more existing independent variables. The word "Logistic" is derived from the Legit function, which is employed in this classification approach

. • Logistic regression is utilised when your Y variable can only take two values, and if the data is linearly separable, it is more efficient to classify it into two distinct groups.

Predictive

Using the categorization algorithms, we can forecast whether or not the IDS attack will occur.

VI. CONCLUSION

We infer that the CSIC_2020 IDS dataset was used as input. The input dataset was described in our study article. We implemented classification methods (i.e., machine learning techniques) such as Logistic Regression.Then, we used an encryption technology like RSA to encode and decode the identified data.Finally, the findings reveal that the accuracy for the aforementioned algorithm and evaluated the performance metrics such as accuracy for algorithms and store the results in the cloud (free storage) for security purposes.

REFERENCES

[1] Y. Liu, X. Y. Ma, L. Shu, G. P. Hancke, and A. M. Abu-Mahfouz, "From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges," IEEE Transactions on Industrial Informatics, vol.17, no.6, pp.4322-4334, June 2021.
[2] G. Aceto, V. Persico, and A. Pescapé, "A survey on information and communication technologies for industry 4.0: state-of-the-art, taxonomies, perspectives, and challenges," IEEE Communications Society Tutorial, vol.21, no.4, pp.3467-3501, August 2019.
[3] Industry 4.0 and cybersecurity: Risk management in the age of linked manufacturing [Online].
[4] O. Friha, M. A. Ferrag, L. Shu, and M. Nafa, "A robust security framework based on blockchain and SDN for fog computing enabled agricultural internet of things," in Proceedings of the International Conference on Internet Things and Intelligent Applications, Zhenjiang, China, 2020, pp. 15.
[5] W. J. Zhu, M. L. Deng, and Q. L. Zhou, "An intrusion detection algorithm for wireless networks based on ASDL,"

ISO 3297:2007 Certified

DOI: 10.17148/IARJSET.2023.10774

IEEE/CAA J. Autom. Sinica, vol.5, no.1, January 2018, pp.92-107.

[6] M. Agarwal, S. Purwar, S. Biswas, and S. Nandi, "Intrusion detection system for PS-poll DoS attack in 802.11 networks using real time discrete event system," IEEE/CAA J. Automobile Sinica, vol. 4, no. 4, pp.792-808, 2017.

[7] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusiondetection: Approaches, datasets, and comparative study," J. Inf. Secur. Appl., vol. 50, p. 102419, February 2020.

[8] "Internet of Things: Wireless Sensor Networks Executive Summary," S. Yinbiao and K. Lee, 2014. E. Cayirci, F. Akyildiz, W. Su, Y. Sankarasubramaniam, and F. Akyildiz

[9] A questionnaire "A survey on sensor networks," IEEE Communications Magazine, vol. 40, no. 8, 2002, pp. 102-105.

[10] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," IEEE Communications Society Tutorials, vol. 11, no. 2, pp. 52-73, 2009.

[11] A.-S. K. Pathan, H.-W. Lee, and C. S. Hong, "Security in Wireless Sensor Networks: Issues and Challenges," 8th Int. Conf. Adv. Commun. Technol., vol. 2, pp. 6–1048, 2006.

[12] E. V. Carrera, A. Gonz'alez, and R. Carrera, "Automated detection of diabeticretinopathy using SVM," 2017 IEEE XXIV International Conference on