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

Suchithra D AMC ENGINEERING COLLEGE

Mrs. Padma Priya AMC ENGINEERING COLLEGE

Abstract

For cultivation 4.0 cyber security, the system, evaluate, and analyse intrusion detection systems. We discuss cyber security risks as fine as assessment criteria that were employed in the performance evaluation of an intrusion detection system for Agriculture 4.0. Then, we assess imposition recognition arrangement in light of upcoming technologies such as cloud computing. disturbance finding is a critical security issue in today's cyber environment. A large range of strategies based on Approaches for deep learning have been devised. As a result, we developed machine learning techniques to identify the invasion. A network-based security system for intrusion detection (NIDS) is a type of invasion detect device. often installed at network points such as gateways and routers to detect network traffic intrusions. We deliver a complete report using the computer's learning methodology. The difficulties and prospective study topics for the future of agriculture computer surveillance are highlighted. To detect dangerous behaviour, intrusion detection systems (IDSs) use artificial intelligence-based technologies including machine training and cloud-based computing. Finally, we may use neural networks for recognising the IDS. and save the observed data in free cloud storage.

KEYWORDS: *IDS, Machine Learning, NIDS*

1. INTRODUCTION

Agriculture 1.0, Agriculture 2.0, Agriculture 3.0, and cultivation 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 generate approving setting for starting point placement and plant growth. Agriculture 2.0 was the name given to a growth in agricultural productivity company was founded at the turn of the twenty-first century agricultural technology such as combines, irrigation, harvesting, trucks, tractors, aeroplanes, helicopters, and so on. Agriculture 3.0 emerged in the early 1970s and is centred on environmentally friendly clean energy sources like bio electricity, thermal energy, and sunlight, and hydropower. The term "Agriculture 4.0" emerged after "Industry 4.0," which is defined by a blend of upcoming technologies that include block chain, software defined networking (SDN), artificial intelligence, Internet of Things (IoT), IoT devices, Helicopters, fog/edge computation, cloud computing, 5G connectivitynetwork function virtualization (NFV), smart grids, and so on. Many new innovations have widely employed in Industry 4.0, and their implementation in agricultural contexts is not difficult to duplicate. As a upshot, the primary difficulty of establishing The future of agriculture is not about deploying new technology, but about ensuring safety and confidentiality, since countless IoT-based sensors are being deployed in a field of grass. Also, there are numerous precautions for safety 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 A network is a collection of systems to detect intrusions (IDS) 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 facts in a free cloud storage service.
- To improve classification algorithms' overall performance.

11. Literature Survey:

2017 study on SDN-based Using systems that detect network intrusions artificial intelligence techniques Methodology: [1] Software Defined Networking (SDN) offers the opportunity to perceive and monitor network security issues caused by the advent of programmable features. to protect networked computers and address network security concerns, Machine Learning (ML) methods have recently been incorporated in SDN-based NIDS (Network Detection and Prevention Systems). The technology known as deep learning (DL) is emerging as a stream of sophisticated algorithmic learning approaches in the framework of SDN. In this work, we assessed several present studies on machine learning (ML) approaches that employ SDN to produce NIDS. Specifically, we evaluated neural network algorithms for buildingSDN-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 several quality culture scheme 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 charge 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. Because the correlation of more than one activity is required to comprehend the assault plan and identify 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 sort of danger and forecast further moves. The object of this revise is to collect all articles that provide multi-step assault detection systems. We concentrate on approaches that Bibliographic research is worn 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 provisos 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-Things devices has increased the volume of real-time network data at a rapid pace. At the equivalent moment, 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 and execution time. Finally, the suggested approach allows analysts to track and spot abnormalities in real time. The spark iterative computation architectural allows large-scale machine education algorithms to reach high levels of efficiency in results, and the spark.ml API for pipeline provides developers with a various place of new modules to interact with their architecture.

• Small and slow-ramped attacks can avoid statistical tactics by limiting The attack's impact fell under scientific thresholds. [4] Using Machine Learning Methods for Detecting Networking Intrusions 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 untimely research in this field and There are many intrusion detection systems on the market (IDS) are signature-based. The issue with signature-based methods is that the database signature must be updated when new attack signatures become They are therefore unsuited for immediate time net identification of anomalies. Automated algorithms for classification have lately gained popularity in the identification of anomalies. On the Kyoto 2006+ data set, we implement and examine seven different machine learning algorithms with entropy of information computing. Our results show that, for Advantage:

• 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 payable to its talent to find and index Internet-connected industrial control equipment. In the job, 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 on industrial control systems in requisites of scanning duration, frequency, scanning port, area preferences, ICS protocol preferences, and ICS protocol function code proportion. As a product, we in attendance several protective methods to lessen the Shodan danger.

The key profit of SVM is that it is a machine learning model through a lofty 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.

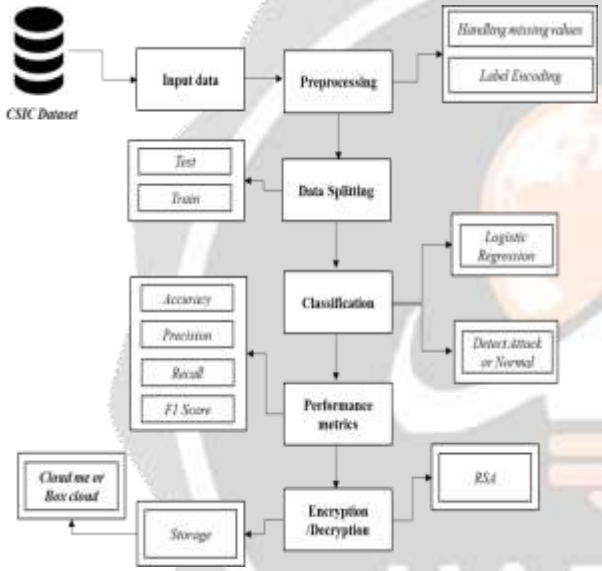

Fig [1] System Architecture

111. 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 Agriculture 4.0 requires a hacking prevention system. After that, we evaluate surveillance technologies. 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. 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.

1V. Proposed System:

The CSIC_2020dataset was used as input in this system. The dataset repository The information that was input was obtained using. This information must then be processed. 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 vast bulk of the information will be available in stream. A smaller portion 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-processing**
- **Data splitting**
- **Classification**
- **Prediction**

[1] Data selection

The The source of the info was obtained. 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 came from somewhere, including the UCI repository.

• The dataset contains data 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.

[2] Pre-processing

Data pre-processing is the process of deleting undesirable data from a dataset.

• Pre-processing data Alteration methods are employed in order to turn the dataset into a machine learning-friendly structure**.** • Washing is also part of this procedur**e.** the dataset by deleting extraneous or damaged data that might impair the dataset's correctness, manufacture it more efficient.

• 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.

• A large number of artificial intelligence methods need quantitative input and output variables.

[3] Data splitting

• Data are required during the machine learning process in order for learning to occur. In addition to the data necessary for education, test data are required to gauge 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 statistics is worn to create a predictive model, while the other is utilised to consider 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

[4] Classification

Techniques for machine learning can be used 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 patchy can only receive two values, and if the data is linearly separable, it is more efficient to classify it into two distinct groups.

[5] Predictive

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

V1. 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 grades 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 contact 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 administration in the age of linked manufacturing [Online].

4. "A strong safety strategy based on blockchain and SDN for cloud computing enabled agriculture web of things," O. Friha, M. A. Ferrag, L. Shu, and M. Nafa, Proceedings of the International Seminar on Internet Things and Intelligent Programmes, Zhenjiang, China, 2020, pp. 15. "An alerting methodology for wireless systems built around ASDL," W. J. Zhu, M. L. Deng, and Q. L. Zhou, IEEE/CAA J. Autom. Sinica, vol.5, no.1, the first month of 2018, pp.92-107.

7. M. Agarwal, S. Purwar, S. Bis were, and S. Nandi, "Intrusion detection approach for PS-poll DoS assaults in 802.11 networking employing real time discrete event system," IEEE/CAA Journal of Automobile Sinica, vol. 4, no. 4, pp.792-808, 2017.

[8] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security incursion being noticed: Approaches, databases, and compared study," J. Inf.

[9] S. Yinbiao and K. Lee, "The internet of Items: Wireless Sensor Networks: An Executive Summary," 2014. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci

10."A survey on sensors connections," IEEE Telecommunications Newspaper, vol. 40, no. 8, 2002, pp. 102-105.

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.