OPEN ACCESS

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 08.05.2023

Accepted:13.06.2023

Published: 01.07.2023

Citation:
Deepa, KR, and A. Rajkumar. "Deep Learning-Based Digital Physical System Threat Identification A Study of Information Security." *Shanlax International Journal of Arts, Science and Humanities*, vol. 11, no. S1, 2023, pp. 249–56.

DOI:
https://doi.org/10.34293/sijash.v11iS1-July.6347

Deep Learning-Based Digital Physical System Threat Identification A Study of Information Security

Deepa K.R
Department of Master of Computer Applications
Rajarajeswari College of Engineering, Bengaluru

A Rajkumar
Department of Master of Computer Applications
Rajarajeswari College of Engineering, Bengaluru

Abstract

With the amplified quantity of cyber-attacks and cyber criminals targeting cyber-physical systems (CPSs), identifying these intrusions remains difficult. Intrusion detection is a critical security issue in today's cyber environment. A large range of strategies based on appliance learning organizations have stood developed. So, in command to sense the infiltration, we created machine learning algorithms. Deep learning (DL) outperforms typical machine learning (ML) methods in terms of performance. When there is enough data, DL models nearly always produce great results. However, as compared to other domains like as NLP, image processing, software vulnerability, and many others, DL models have been slowly deployed to attack the CPS cybersecurity issue.

Many DL mock upstake likewise stood accessible in recent articles to identify CPS cyber-attacks. A commonly acknowledged explanation for the problems in identifying cyber-attacks on CPSs is the gradation of complexity when superimposing cybersecurity over CPSs. The dataset UNSW-NB15 was used in this system from the dataset repository. Then we must implement several classification methods such as logistic regression (LR) and LSTM. The untried findings recommend that equally methods are accurate.

Keywords: Cyber-Physical System, Cybersecurity, Deep Learning, Intrusion Detection, Pattern Classification.

Introduction

When cyber-physical systems (CPSs) are increasingly connected to the cyber environment, they are vulnerable to cyber-attacks. Cyber-attacks have gotten more sophisticated and common as automated assaulting tools have been available, and professional hacker groups have begun to participate. A successful cyber assault on a CPS might be devastating, catastrophic, or even lethal.

Because many CPS systems lack cyber security safeguards such as message authentication, it is problematic to recognize fake data injection attacks. A lack of widespread encryption, particularly on systems using outdated technology, styles it problematic to defend against eavesdropping assaults.

As technology advances, the quantity of hacking occurrences rises. Companies report a huge number of hacking instances each year. In 2007, a disseminateddisavowal of package assault was attempted against Estonian websites.Amazon began receiving authenticated requests from numerous users at one of their locations on June 17, 2008.

The websites had to slow down as a consequence of a sharp rise in the quantity of queries. The Mediafire service was significantly interrupted for a period of over fifteen hours on January 13, 2013, according to an announcement made by the European Networking and Data Protection Directorate. This disruption affected all users globally. the social network was allegedly the victim of an international denial if service occurrence on September 29, 2014. 50% of computer hacking, it has been reported, start with some sort of internet surveillance operation.

Hackers risk compromising sensitive data stored on devices by not only executing inundation and penetrating attacks but also by propagating malicious programmes in the manner of Trojan horses, worms, and spams. 40% of all electronic mail sent internationally on April 18, 2013, were related to the Boston Boston bombardment, according to the Cisco Corporation Quarterly Safety Report. Torpedo constitutes one of the top ten malware used to get preliminary entry into user PCs and corporate networks, according to a 2017 Cisco investigation.

As a result, Privacy is a main worry that wants to be managed carefully in a highly sophisticated technological setting. For detecting intrusions, researchers investigated a separate group of assaults. For instance, founded on the KDD'99 dataset, Attacking types include Relational to Locally (R2L), Tenant to Rooted (U2R), and Disruption of Service (DoS) (Bandwidth and Capacity Depletion) threats. Processes are alienatedhooked on nines distinct groups pursuant to a contemporary threat datasets (UNSW-NB): Fuzzer, Assessment, Investigation, ShellCode, Worm, Generic, DoS, Exploit, and General.

Segment III driveshooked oncountlesscomplexity on respectively of these attacks. Middle-boxes like as firewalls, antivirus, and impositionuncovering systems (IDS) are being used in security solutions. A firewall regulates network traffic depending on the foundation or terminus address. It modifies traffic in accordance with the firewall instructions. Firewalls are likewise constrained by the amount of national they have accessible as well as their information of the crowdsgetting the material. An impositionuncoveringorganization (IDS) is a sort of safetyinstrument that analyses system traffic and images the system for distrustful activity before alerting the organization or network administrator.

A network-based impositionuncovering system (NIDS) is often installed at system points such as doorways and routerss to detect network traffic intrusions. These IDSes utilise three sorts of uncoveringapparatuses at the highest level: abuse detection, incongruity detection, and hybrid detection. In the misuse detection strategy, the IDS keeps a set of informationcenters (rules) for recognizing known attack types.

Misapplicationuncovering systems are widely classed as knowledge-based or machine learning-based. The knowledge-based approach compares systemcirculation or System calls belong traces, for example) to established standards or methods of attack. Knowledge-based methodsdecreasehooked on three chief groups: Analysing state transitions and (i) matching signatures(ii), and rule-based expert systems (iii).

Several articles looked into data-driven strategies for identifying cyber-attacks on CPS systems. There is, however, no extensive discussion of using DL approaches to identify CPS cyber-attacks. Without a special focus on cybersecurity, a short survey was supplied with a four-step framework for applying DL approaches to CPS challenges like as cybersecurity, adaptability, recoverability, and many more.

Without looking into the DL models, a complete study of cyber-attacks against CPSs was published in. Without employing DL techniques, many approaches of detecting cyber-attacks in CPSs were summarised in. A complete list of CPS attacks and obstacles was published in, although ML and DL techniques were left out.

Objective of the Work

The primary goal of our study is to properly categorise or forecast cyber-attacks in networks. To improve efficiency, implement several categorization techniques.
To improve classification algorithms' performance as a whole.

Literature Survey

Nasrin Sultana[1] claimed that Software DistinctSchmoozing Technology (SDN) offers the opening to perceive and monitor network security issues due to 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 System Intrusion Detection Systems (NIDS). In the background of SDN, a stream of advanced machine learning methodologies - deep learning knowledge (DL) - is beginning to develop. We evaluated different current research on machine learning (ML) methodologies that use SDN to create NIDS in this schoolwork.

We especially studied deep learning approaches in the expansion of SDN-based NIDS. Temporarily, 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 features knowledgemeans are their complexity and high implementation costs.

Julio Navarro[2] talked about it. 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 command to achieve their ultimate goal, attackers must now go through many intrusive procedures. The collection of these processes is denoted to as a multi-step assault, multi-stage attack, or attack scenario. Because the correlation of more than one activity is compulsory 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 type of danger and forecast further moves.

The box of this schoolwork is to collect all articles that provide multi-step assault detection systems. We concentrate on approaches that go beyond detecting a symptom and attempt to expose the entire structure of the attack in addition to the associations between its phases. To locate relevant material, we use a methodical approach to bibliographic research. 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 organization is that it detects harmful network events using IDS signatures and tracks their progression as successive events, looking for matches in rapports of IP address or port.
- Because an attacker does not necessity to follow a certain order when performing a multi-step assault, the collection of alternative action sequences might be extremely complicated.

According to Riyaz Ahmed[3], the ever-increasing use of linked Internet-of-Things devices has recently increased the volume of real-time network data with high velocity. 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, phantomgathering, and agglomerative clustering are secondhand to undertake critical comparative analysis. When compared to other algorithms, the evaluation results demonstrated the usefulness of the suggested outline with a considerably healthiercorrectness rate of 96.51%.

- The spark iterative computation architectural allows large-scale machine educationprocesses to reach high levels of efficiency in results, and the spark.ml API for cylinder provides designers with a diverse set of new components to interact with their architecture.
- Small and slow-ramped attacks can avoid statistical tactics by limiting the impact of the attack below statistical criteria.

According to Marzia Zaman[4], network traffic anomalies might suggest a probable network breach, hence anomaly uncovering is critical for detecting and preventing security assaults. The popular of the early research in this arena and commercially accessibleImpositionUncovering Systems (IDS) are signature-based. The issue with signature-based methods is that the catalogueautograph must be updated when new attack signatures become accessible, creation them unbecoming for actual-periodsystemincongruity detection. Machine learning cataloguingapproaches have latelydevelopedgeneral in anomaly detection.

We implement and analyse seven substitute machine learning methods with information entropy computation to the Kyoto 2006+ data set. Our conclusions reveal that maximum mechanism education procedures deliver greater than 90% precision, recall, and accurateness for this particular data set. However, using the area under the Receiver Operating Curve (ROC) measure, we discover that the Circular Basis Function (RBF) outperforms the other seven methods investigated in this paper.

- There is little training time.
- The fundamental disadvantage of this autograph-founded technique is that the database autograph must be updated when new autographs become available, making it unsuitable for real-time system anomaly uncovering.
- Comparing the outcomes of the sevens methods mentioned here using numerous performance indicators is quite tricky.

Dan Yu[5] talks about The Industrial Control System (ICS), as a crucial component of critical infrastructure, is increasingly vulnerable to cyber assaults. The appearance of the Shodanexplorationmachine heightened the threat. The Shodan search engine has become a favourite toolbox for aggressors 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. We create a graded DFA-SVM identification model to detect Shodan X-raysgrounded on function code and traffic characteristic, which is customised to trace Shodan and Shodan-like scanners superior to the prevailing approach of reverse resolving IPs.

Finally, we undertake an in-depth study of Shodan X-rays and assess the influence of Shodan on industrial control organizations in rapports of scanning duration, frequency, scanning port, area preferences, ICS custom preferences, and fraction of ICS etiquette function codes. As a outcome, we current several protective methods to lessen the Shodan danger.

- The key benefit of SVM is that it is a machine learning model with a high recognition rate of tiny samples and a good generalisation ability, making it suited for handling high-dimensional and non-linear Shodan traffic from a limited quantity of Shodan scanners.
- If the entireamount of data is large, the training time is long.

- The key advantage is that heterogeneity across real Sensors, IDSs, Analyzers, or even SIEMs can be useful for Intrusion Detection, since detection accuracy can be enhanced.

Existing Model

In the present system, a comprehensive perspective of newly proposed DL techniques for cyber-attack detection in the CPS context is offered. To summarise and analyse the examined literature for applying DL approaches to identify cyber-attacks on CPS systems, a six-step DL-driven methodology is offered. CPS scenario analysis, cyber-attack identification, ML issue formulation, DL model modification, data collecting for training, and presentation assessment are all helping of the technique. The examined studies show that DL modules have a high potential for detecting cyber-attacks on CPS. Furthermore, outstanding performance is accomplished in part because to the availability of various high-quality datasets for public usage. Furthermore, future research difficulties, opportunities, and research trends are identified.

- It is inefficient for big amounts of data.
- It did not use the machine learning algorithm.
- Theoretical bounds.
- The procedure is carried out without the removal of unnecessary data.

Proposed Methodology

The UNSW-NB15 datasets was used as input in this system. The dataset source was used to obtain the effort 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 separated into two parts: test and train. The datas is being separated depending on a ratio. The popular of the data will be present in train. A reduced fraction of the data will be contemporary in the exam. The drillchapter is used to measure the model, whereas the testing phase is secondhand to forecast the model.The classification algorithm (i.e., machine and deep learning) must next be implemented. Logistic regression and deep education algorithms are instances of machine learning algorithms.

LSTM, for example. Finally, the experimental falloutsvalidate that presentation measurements like accuracy and comparative outcomes are important.

- It is efficient for huge datasets;
- It consumes less time; and
- It is accomplished by deleting unnecessary data.

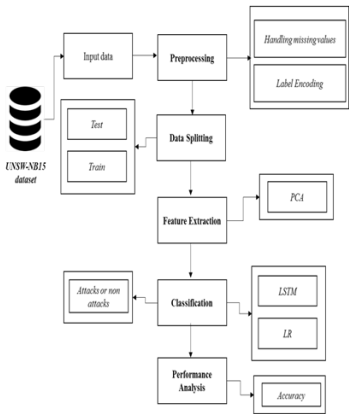

Figure 1 Proposed Architecture

Implementations

Data Selection

The information that was input was gainedsince a dataset repository, and the UNSW-NB15 dataset was employed in our procedure.The process of identifying cyber-attacks begins with data selection.The input datasets was gotten from a source such as the UCI fountain.

The dataset includes data such as protocol, duration, host error rate, attack category, sticker, and so on. We can read or load our input datasets in Python using the panda module.Our dataset is in the '.csv' format.

Data Preprocessing

In this module, we develop the Doctor's part, where the new doctor is registered by entering their facts in the registration form. Once after registration the doctor cannot able to login in to the system comparable to the earlier module. Only if the cloud server approves the doctor only they can login into the system, this is developed to make the system more secure. The doctor module provides authorized doctors with access to patients' PHRs. It allows them to search for patients available securely and ensure the confidentiality of the PHRs.

Data Splitting

Data are required throughout the machine learning process in order for learning to occur.
In calculation to the statistics necessary for training, test data are compulsory to assess the algorithm's performance and determine how effectively it performs.

We regarded 80% of the participation dataset to be training data and the additional 20% to be testing data in our procedure.

The development of dividing accessible datas into two sections, commonly for irritable-validators reasons, is identified as datas splitting.

One helping of the information is used to create a predictive model, while the other is utilised to evaluate the replica'spresentation.

It is dangerous to separate data into exercise and difficult sets when analysing data mining methods.

Normally, when you division a data collection into.

Feature Classification

Data are required throughout 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 80% of the input dataset to be training data and the other 20% to be testing data in our procedure.

The course of dividing accessible data into two sections, commonly for cross-validator reasons, is known as data splitting.

One portion of the data is hand-me-down to create a predictive model, while the other is utilised to assess the model's performance.

It is critical to separate data into exercise and difficult sets when analysing data mining methods. Normally, when you division a statistics collection into.

Classification

In our approach, we must use numerouscataloguingprocesses such as LR and LSTM.

Logistic Regression is a Machine Learning method that is secondhand for classification issues; it is a predictive analytic approach that is grounded on the probability notion.

The logistic regression hypothesis tends to restrict the cost function between 0 and 1.

Result

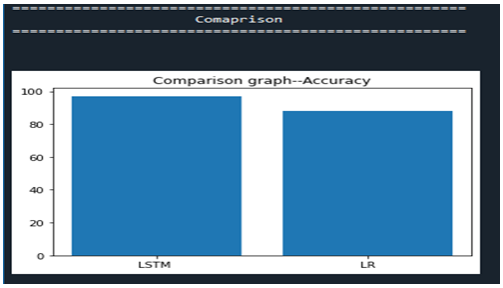

Figure 1

Fig 1 shows The Final Result will get generated based on the overall classification and prediction.

Conclusions

As a outcome, we infer that the UNSW-NB15 dataset was used as input. Our study paper highlighted the input dataset. We used machine and deep learning methods to develop classification processes. Then there are machine learning algorithms like logistic regression and deep learning algorithms like LSTM. Finally, the output demonstrates the accuracy for the aforementioned algorithm as well as estimated performance measures such as accuracy for both methods and comparison graph.

Future Work

In the future, we'd like to combine two separate machine learning algorithms or two deep learning algorithms. It is feasible to give enhancements or modifications to the suggested clustering and classification algorithms in the future to obtain even higher performance. Aside from the tried-and-true combination of data withdrawal techniques, additional combinations and clustering algorithms can be utilise to increase detection accuracy.

References

1. S. Yinbiao and K. Lee, "Internet of Things: Wireless Sensor Networks Executive summary," 2014. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci
2. "A survey on sensor networks," IEEE Commun. Mag., vol. 40, no. 8, pp. 102–105, 2002.
3. X. Chenu, K. Makkin, K. Yenoij, and N. Pissinouij, "IEEE Communications Society Instructions, "Sensor networking securites: a survey," vol. 11, no. 2, pp. 52–73, 2009.
4. A.-S. K. Pathann, H.-W. Leeeg, and C. S. Hongg, ""Problems and issues with security in wireless sensor networking." 8th International Conference on Advanced Communications Technology, 2006, vol. 2, pp. 6-1048.
5. P. Yi, Y. Jiange, Y. Zhongo, and S. Zhangk, "Distribute Intrusion Detection for Mobile Ad Hoc Networks," 2005 Symp. Applied Online Working (SAINT 2005 Work), pp. 94–97.
6. H. Sedjelmaci and M. Fehamy, "Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network," International Journal of Network Security Applications, Volume 3, Number 4, July 2011, Pages 1–14.
7. L. Khann, M. Awadi, and B. Thuraisinghamuy, "Support vector machines and hierarchical clustering are used in a new intrusion detection system, VLDB J., vol. 16, no. 4, pp. 507-521, 2007.
8. S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014, pp. 1348–1353, 2014.

9. O. Can, C. Turguner, and O. K. Sahingoz, "A Neural Network Based Intrusion Detection System For Wireless Sensor Networks," Signal Process. Commun. Appl. Conf. (SIU), 2015 23th, pp. 2302–2305, 2015.
10. F. Lu and L. Wang, "Intrusion Detection System Based on Integration of Neural Network for Wireless Sensor Network," J. Softw. Eng. 2014.
11. Y. Y. Li and L. E. Parker, "Intruder detection using a wireless sensor network with an intelligent mobile robot response," Southeastcon, 2008. IEEE, pp. 37–42, 2008.
12. A. Kulakov and D. Davcev, "Tracking of unusual events in wireless sensor networks based on artificial neural-networks algorithms," Inf. Technol. Coding Comput. 2005. ITCC 2005. Int. Conf., pp. 534–539, 2005.
13. M. Panda, "Security Threats at Each Layer of Wireless Sensor Networks," Int. J. Adv. Res. Comput.Sci. Softw. Eng.
14. Karlof and D. Wagner, "Secure routing in wireless sensor networks: attacks and countermeasures," Proc. First IEEE Int. Work. Sens. Netw. Protoc. Appl. 2003., pp. 113–127, 2003.
15. H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson,and X. Bellekens, "A Taxonomy and Survey of Intrusion DetectionSystem Design Techniques, Network Threats and Datasets," vol. 1,no. 1, 2018.
16. J. Navarro, A. Deruyver, and P. Parrend, "A systematic survey on multi-step attack detection," Computers & Security, vol. 76, pp. 214–249, 2018.
17. R. Zuech, T. M. Khoshgoftaar, and R. Wald, "Intrusion detection and Big Heterogeneous Data: a Survey," Journal of Big Data.
18. J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation," Proceedings of the 1st Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2011, pp. 29–36, 2011.
19. K. Kendall and A. C. Smith, "A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems by A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems," 1999.