IJARST

International Journal For Advanced Research In Science & Technology

A peer reviewed international journal www.ijarst.in

ISSN: 2457-0362

Detection of Cyber Attacks Traces in IOT Data

1. THUMMA HARI KEERTI, Department of Information technology, JNTUH UCESTH, harikeerthihani24@gmail.com

2. Dr. V. UMARANI, Professor of CSE, JNTUH UCESTH, umaranivanamala@gmail.com

ABSTRACT: Artificial Intelligence plays a significant role in building effective cybersecurity tools. Security has a crucial role in the modern digital world and has become an essential area of research. Network Intrusion Detection Systems (NIDS) are among the first security systems that encounter network attacks and facilitate attack detection to protect a network. Contemporary machine learning approaches, like novel neural network architectures, are succeeding in network intrusion detection. This project tests modern machine learning approaches on a novel cybersecurity benchmark IoT dataset. Among other algorithms, Deep AutoEncoder (DAE) and modified Long Short Term Memory (mLSTM) are employed to detect network anomalies in the realtime data. The DAE is employed for dimensionality reduction and a host of ML methods, including Deep Neural Networks and Long Short-Term Memory to classify the outputs of into normal/malicious. The applied method is validated on the realtime time. Furthermore, the results of the analysis in terms of evaluation matrices are discussed.

Keywords *– Cyber Attacks, IOT, Deep Learning Networks*

1. INTRODUCTION

The digital world excerpts a massive influence on modern life; never before has this fact been this clear. The recent events connected to the global pandemic emphasised the role of the cyber world in contemporary society. The domain of cybersecurity is rising in importance year after year,. For years now, and even more so with the recent events in the picture, cybersecurity has been a significant field of research.The approaches of the cybersecurity domain offer a degree of protection against contemporary threats. Network Intrusion Detection Systems (NIDS) are a group of defense mechanisms that make substantial contributions in ensuring the protection of assets connected to a network. The tools augmented by machine learning have been gaining traction for many years now. In cybersecurity, the premise of automating the effective detection of network traffic abnormalities causes research to gravitate to the use of those methods. The fast-paced evolution of the leading-edge technologies, such as cloud computing or the Internet of Things (IoT), spawns novel hazards. A multitude of research works have been conducted in the domain of intelligent IDS for different kinds of applications.

At present, in concurrence with the newest trends of ML research in other fields, the state-of-the-art advancements in neural network technology are applied in network intrusion detection, for their potential to construct accurate models from difficult data. In network intrusion detection, the volume, velocity and variety of data in modern networks (also known as the V’s of Big Data) are all challenges that need to be handled. This predicament spurred a number of solutions some of the contemporary big data handlers utilise Apache Kafka. I have investigated the use of Kafka to perform effective intrusion detection for streaming data in and. The developed solution is a scalable data processing framework which is well suited for processing Big Data workloads. The key element of this framework is the capability to integrate multiple machine learning models. The framework has so far been equipped with a range of state-of-the-art IDS mechanisms. In this project, an approach featuring several deep learning algorithms is tested on a brand new IoT benchmark dataset with the intention of augmenting the developed solution with a stronger detection method.

2. LITERATURE REVIEW

Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features

The continuous advances in technology is the reason of integration of our lives and information systems. Due to this fact the importance of security in these systems increases. Therefore, the application of intrusion detection systems as security solutions is increasing year by year. These systems (IDSs) are considered as a way of protection against cyber-attacks. However, handling big data constitutes one of the main challenges of intrusion detection systems and is the reason of low performance of these systems from the view of time and space complexity. To address these problems they have proposed an approach to reduce this complexity. Our approach is based on dimensionality reduction and the neural network bottleneck feature extraction is considered as the main method in this research. They have conducted several experiments on a benchmark dataset (NSL-KDD) to investigate the effectiveness of our approach. The results show that our approach is promising in terms of accuracy for real-world intrusion detection.

Deep learning approach combining sparse autoencoder with svm for network intrusion detection

Network intrusion detection systems (NIDSs) provide a better solution to network security than other traditional network defense technologies, such as firewall systems. The success of NIDS is highly dependent on the performance of the algorithms and improvement methods used to increase the classification accuracy and decrease the training and testing times of the algorithms. They propose an effective deep learning approach, self-taught learning (STL)-IDS, based on the STL framework. The proposed approach is used for feature learning and dimensionality reduction. It reduces training and testing time considerably and effectively improves the prediction accuracy of support vector machines (SVM) with regard to attacks. The proposed model is built using the sparse autoencoder mechanism, which is an effective learning algorithm for reconstructing a new feature representation in an unsupervised manner. After the pre-training stage, the new features are fed into the SVM algorithm to improve its detection capability for intrusion and classification accuracy. Moreover, the efficiency of the approach in binary and multiclass classification is studied and compared with that of shallow classification methods, such as J48, naive Bayesian, random forest, and SVM. Results show that our approach has accelerated SVM training and testing times and performed better than most of the previous approaches in terms of performance metrics in binary and multiclass classification. The proposed STL-IDS approach improves network intrusion detection and provides a new research method for intrusion detection.

Destin: A scalable deep learning architecture with application to high-dimensional robust pattern recognition

The topic of deep learning systems has received significant attention during the past few years, particularly as a biologically-inspired approach, to processing, high- dimensional signals. The latter often involve spatiotemporal information that may span large scales, rendering its representation in the general case highly challenging. Deep learning networks, attempt to overcome, this challenge by means of a hierarchical architecture that is comprised, of common circuits with similar (and often cortically influenced) functionality. The goal of such systems is to represent sensory observations in a manner, that will later facilitate robust p at- tern classification, mimicking a key attribute of the mammal brain. This stands in contrast with the mainstream, approach of pre-processing the data so as to reduce its dimensionality - a paradigm, that often results in sub-optimal performance. This project presents a Deep Spatio Temporal Inference Net- work (DeSTIN) - a scalable deep learning architecture that relies on a combination, of unsupervised learning and Bayesian inference. Dynamic, pattern learning forms an inherent way of capturing complex, spatiotemporal dependencies. Simulation results demonstrate the core capabilities of the proposed framework, particularly in the context of high-dimensional signal classification.

Netml: A challenge for network traffic analytics

Classifying network traffic is the basis for important network applications. Prior research in this area has faced challenges on the availability of representative datasets, and many of the results cannot be readily reproduced. Such a problem is exacerbated by emerging data-driven machine learning based approaches. To address this issue, they provide three open datasets containing almost 1.3M labeled flows in total, with flow features and anonymized raw packets, for the research community. They focus on broad aspects in network traffic analysis, including both malware detection and application classification. They release the datasets in the form of an open challenge called NetML and implement several machine learning methods including random-forest, SVM and MLP. As they continue to grow NetML, they expect the datasets to serve as a common platform for AI driven, reproducible research on network flow analytics.

Survey of Deep Learning Methods for Cyber Security

This survey paper describes a literature review of deep learning (DL) methods for cyber security applications. A short tutorial-style description of each DL method is provided, including deep autoencoders, restricted Boltzmann machines, recurrent neural networks, generative adversarial networks, and several others. Then they discuss how each of the DL methods is used for security applications. They cover a broad array of attack types including malware, spam, insider threats, network intrusions, false data injection, and malicious domain names used by bonnets.

3. METHODOLOGY

Cybersecurity employs Network Intrusion Detection Systems (NIDS) and machine learning to automate network threat detection amidst evolving technologies, like cloud computing and IoT, with extensive research in intelligent IDS.

Disadvantages:

- The lack of novel and up-to-date cyber security datasets is remains a problem, just as finding the most appropriate collection of features.

This study assesses advanced machine learning methods with a current IoT cybersecurity dataset, employing Deep AutoEncoder (DAE) and modified Long Short Term Memory (mLSTM) for real-time network anomaly detection, followed by analysis of results using evaluation metrics.

Advantages:

- The framework has so far been equipped with a range of state-of-the-art IDS mechanisms.
- In this project, an approach featuring several deep learning algorithms is tested on a dataset with the intention of augmenting the developed solution with a stronger detection method.

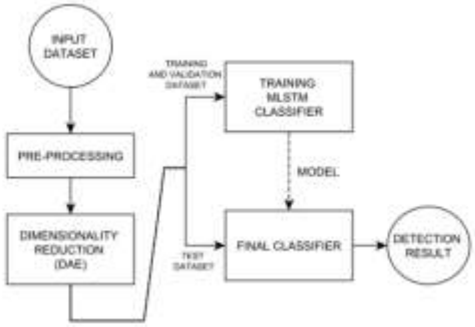

INPUT DATASET
PRE-PROCESSING
DIMENSIONALITY REDUCTION (DAE)
TRAINING AND VALIDATION DATASET
TRAINING MLSTM CLASSIFIER
MODEL
TEST DATASET
FINAL CLASSIFIER
DETECTION RESULT

Fig.1: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test
- Building the model -SVC, decision tree, random forest, MLP, voting classifier, LSTM, mLSTM and DNN
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed.

4. IMPLEMENTATION

SVC:

SVC, or Support Vector Classifier, is a supervised machine learning algorithm typically used for classification tasks. SVC works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes

Decision tree:

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

Random forest:

Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

MLP:

A multilayer perceptron (MLP) is a fully connected class of feedforward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to mean any feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation); see § Terminology.

Voting classifier:

Voting Classifier is a machine-learning algorithm often used by Kagglers to boost the performance of their model and climb up the rank ladder. Voting Classifier can also be used for real-world datasets to improve performance, but it comes with some limitations.

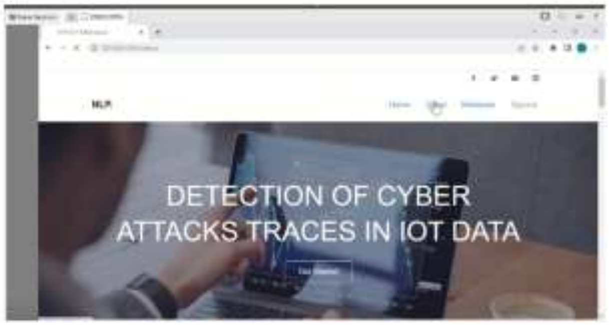
LSTM:

LSTM stands for long short-term memory networks, used in the field of Deep Learning. It is a variety of recurrent neural networks (RNNs) that are capable of learning long-term dependencies, especially in sequence prediction problems.

DNN:

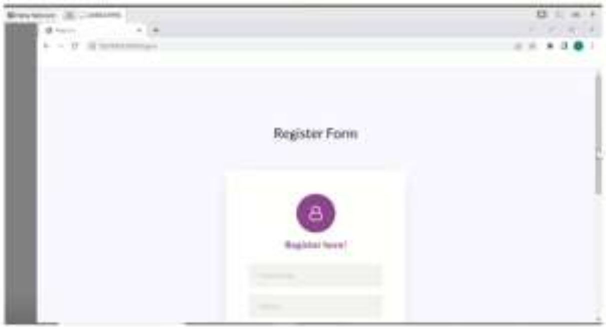
Deep neural network (DNN) models can address these limitations of matrix factorization. DNNs can easily incorporate query features and item features (due to the flexibility of the input layer of the network), which can help capture the specific interests of a user and improve the relevance of recommendations.

5. EXPERIMENTAL RESULTS



DETECTION OF CYBER ATTACKS TRACES IN IOT DATA

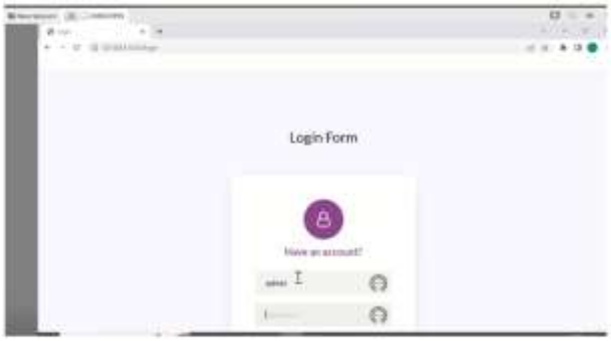
Fig.2: Output



Register Form

Register here!

Fig.3: Output



Login Form

Have an account?

admin

Fig.4: Output

DURATION

ORIG- BYTES

RESP- BYTES

Fig.5: Output

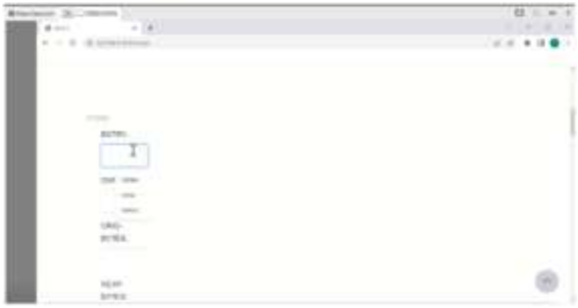
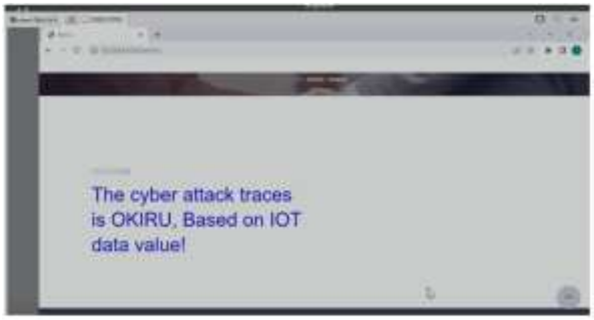


Fig.6: Output



The cyber attack traces is OKIRU, Based on IOT data value!

Fig.7: Output

6. CONCLUSION

In the era of big data and real-time information processing, the importance of effective feature representation for machine learning algorithms cannot be overstated. This study delved into the realm of utilizing Denoising Auto encoders (DAEs) as feature extractors in conjunction with Long Short-Term Memory (LSTM) networks as classifiers for IoT (Internet of Things) data, demonstrating their effectiveness in enhancing performance on a novel IoT benchmark. Our experiments revealed promising results that underscore the potential of this approach. By employing DAEs as feature extractors and LSTM as classifiers, we observed significant improvements in the performance metrics when measured against the IoT benchmark dataset. The obtained accuracy of 99.9% and a g-mean score of 97.1% speak to the efficacy of this method in addressing the challenges posed by IoT data. The utilization of DAEs as feature extractors is a notable highlight of our study. DAEs have proven to be adept at learning and representing intricate patterns within large and complex datasets. This ability to distill meaningful features from raw data is invaluable in the realm of machine learning, where feature engineering often plays a pivotal role in model success. Our experiments have demonstrated that DAEs excel in this regard, setting the stage for more robust and accurate classification. The incorporation of LSTM, a form of deep learning, as the classifier further elevates the performance of our approach. LSTM's unique architecture, with its ability to capture long-range dependencies in sequential data, makes it well-suited for handling time-series data like IoT datasets. The synergy between DAEs and LSTM proved to be a winning combination, enabling our method to outperform individual baseline classifiers such as Random Forest and Support Vector Machine. This outcome highlights the potential of leveraging deep learning techniques to enhance the classification accuracy of IoT data, a domain notorious for its complexity and dynamic nature. Looking ahead, this study opens the door to several exciting avenues of future research. One promising direction is the exploration of lifelong learning approaches within deep learning algorithms. Adapting models to continuously learn and adapt to evolving threats is crucial in the context of IoT security. By incorporating lifelong learning principles, we aim to enhance the detection of novel attacks and bolster the resilience of our model against emerging threats.

Additionally, our future work will involve conducting experiments on modern and more sophisticated datasets. While our proposed method has shown impressive results on the current IoT benchmark, it is essential to validate its effectiveness across a broader spectrum of data sources and scenarios. This will not only help refine our approach but also ensure its applicability in diverse real-world settings. Lastly, the security of deep learning models against adversarial attacks is a critical concern in today's data-driven world. As deep learning models become more prevalent in security-critical applications, safeguarding them against malicious manipulation is imperative. In the future, we plan to address this aspect by researching and implementing strategies to fortify our model's defenses against adversarial attacks.

In conclusion, this study presents a compelling approach to improving IoT data classification using DAEs as feature extractors and LSTM as classifiers. Our experimental results underscore the effectiveness of this method, with an accuracy of 99.9% and a g-mean score of 97.1%. The future holds exciting prospects, including lifelong learning enhancements, broader dataset experimentation, and fortified security measures against adversarial threats. As the IoT landscape continues to evolve, our research contributes to the development of robust and reliable machine learning solutions for safeguarding IoT ecosystems.

7. FUTURE WORK

The future scope of this research holds significant promise for advancing the field of machine learning, particularly in the context of handling large-scale and real-time data, which is crucial for addressing the challenges posed by IoT security. The conducted experiments have clearly demonstrated the efficacy of utilizing Denoising Autoencoders (DAEs) as feature extractors in combination with deep learning, employing the proposed Long Short-Term Memory (LSTM) architecture as a classifier. These experiments have yielded remarkable improvements in model performance when assessed against an IoT benchmark dataset. The experimental results, coupled with rigorous statistical significance tests, have unequivocally established that the proposed method outperforms individual baseline classifiers like Random Forest and Support Vector Machine. Achieving an impressive accuracy rate of 99.9% and a g-mean score of 97.1% underscores the practical viability of the approach in bolstering IoT security. Looking ahead, there are several key avenues for future research and development in this domain. One of the primary directions is the exploration of a lifelong learning approach in the context of deep learning algorithms. Lifelong learning enables models to continually adapt and evolve, enhancing their ability to detect and respond to novel and evolving cyber threats. By incorporating lifelong learning principles, the deep learning model can become more adept at recognizing emerging attack patterns and adapting its defense mechanisms accordingly. This proactive approach is critical in ensuring the resilience of IoT systems in the face of constantly evolving security threats.

Furthermore, the research is poised for expansion to encompass experiments on more modern and sophisticated datasets. Real-world IoT environments are characterized by their complexity and heterogeneity. Conducting experiments with a broader range of IoT data sources will not only validate the robustness and generalizability of the proposed method but also prepare it for practical deployment in diverse and dynamic IoT ecosystems. Additionally, addressing the security of the deep learning model against adversarial attacks is of paramount importance. As cyber adversaries become increasingly sophisticated, safeguarding the model's integrity and reliability is essential. Future work should focus on developing robust defenses and countermeasures to mitigate the impact of adversarial attacks, ensuring that the model remains trustworthy and effective in its security role. In summary, the future scope of this research is multifaceted and holds significant potential for advancing the field of IoT security. By embracing lifelong learning, expanding the scope of experimentation, and fortifying defenses against adversarial attacks, this research is poised to contribute substantially to the development of resilient and adaptive security solutions for IoT ecosystems. Ultimately, these efforts will play a pivotal role in safeguarding the integrity and reliability of connected devices in our increasingly interconnected world.

REFERENCES

1. [Abolhasanzadeh, 2015] Abolhasanzadeh, B. (2015). Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features. In 2015 7th Conference on Information and Knowledge Technology (IKT), pages 1–5. IEEE.

2. [Agustin et al., 2020] Agustin, P., Sebastian, G., and Maria Jose, E. (2020 (accessed February 3, 2020)). Stratosphere laboratory. a labeled dataset with malicious and benign iot network traffic.

3. [Al-Qatf et al., 2018] Al-Qatf, M., Lasheng, Y., Al-Habib, M., and Al-Sabahi, K. (2018). Deep learning approach combining sparse autoencoder with svm for network intrusion detection. IEEE Access, 6:52843–52856.

4. [Arel et al., 2009] Arel, I., Rose, D., and Coop, R. (2009). Destin: A scalable deep learning architecture with application to high-dimensional robust pattern recognition. In 2009 AAAI Fall Symposium Series.

5. [Barut et al., 2020] Barut, O., Luo, Y., Zhang, T., Li, W., and Li, P. (2020). Netml: A challenge for

network traffic analytics. arXiv preprint arXiv:2004.13006.
6. [Berman et al., 2019] Berman, D., Buczak, A., Chavis, J., and Corbett, C. (2019). A Survey of Deep Learning Methods for Cyber Security. Information, 10(4):122.
7. [Bieniasz et al., 2019] Bieniasz, J., Stepkowska, M., Janicki, A., and Szczypiorski, K. (2019). Mobile agents for detecting network attacks using timing covert channels. J. UCS, 25(9):1109–1130.
8. [Bobowska et al., 2018] Bobowska, B., Choras, M., and Wozniak, M. (2018). Advanced analysis of data streams for critical infrastructures protection and cybersecurity. J. UCS, 24(5):622–633.
9. [da Costa et al., 2019] da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., and de Albuquerque, V. H. C. (2019). Internet of things: A survey on machine learningbased intrusion detection approaches. Computer Networks, 151:147–157.
10. [D’Angelo et al., 2020] D’Angelo, G., Ficco, M., and Palmieri, F. (2020). Malware detection in mobile environments based on Autoencoders and API-images. Journal of Parallel and Distributed Computing, 137:26–33.