Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 14, No. 2, No.01 (2023).
ISSN : **1906-9685**

Journal of Nonlinear Analysis and Optimization : Theory & Applications

Editors-in-Chief :

AN IOT BASED MACHINE LEARNING TECHNIQUE FOR EFFICIENT ONLINE LOAD FORECASTING

R.YAMUNA Assistant Professor, Department of Electrical and ElectronicsEngineering ,VSB College of Engineering Technical Campus,Coimbatore.Email Id : yamunarangaraj@gmail.com
D.PRIYADHARSHINI Assistant Professor, Department of Electrical and Electronics Engineering, VSB College of Engineering Technical Campus, Coimbatore. Email Id: drpriyadharshini.cbe@gmail.com
N.HEMALATHA Assistant Professor, Department of Electrical and Electronics Engineering, Pollachi Institute of Engineering and Technology, Pollachi. Email Id: priyahems25@gmail.com

ABSTRACT

Internet of Things (IoT) networks are computer networks that have an extreme issue with IT security and an issue with the monitoring of computer threats in specific. The paper proposes a combination of machine learning methods and parallel data analysis to address this challenge. The architecture and a new approach to the combination of the key classifiers intended for IoT network attacks are being developed. The issue classification statement is created in which the consistency ratio to training time is the integral measure of effectiveness. To improve the preparation and assessment pace, it is suggested to use the data processing and multi-threaded mode offered by Spark. In comparison, a preprocessing data set approach is proposed, resulting in a significant reduction in the length of the sample. An experimental review of the proposed approach reveals that the precision of IoT network attack detection is 100%, and the processing speed of the data collection increases with the number of parallel threads.

Keywords: Machine learning, parallel processing, classifier design and evaluation

1. INTRODUCTION

The concept oftheInternet of Things (IoT) is increasingly used in many areas of computer and technology engineering such as hospitals, residential administration, automation, traffic control, transport by rail and air, energy usage, development, etc. There is an apparent growth of IoT networks. IoT networks have key characteristics that separate them from many traditional computing networks. They link traditional computer network nodes, electronic "things," all kinds of devices and networks (Wi-Fi, sensors, cell networks, and other networks) into a shared global computer network. They are, therefore, linked to different types of networks. It ensures the interchange of products and conventional computing equipment. Other characteristics of IoT include high heterogeneousness of its components, geographic distribution and very complex physical and logical structure with multiple interface points, low network node computation capacities and high data flow and transmission volumes. The features of IoT networks are very acute in security. The use of proven information protection resources which are adequate for conventional computer networks in IoT networks becomes ineffective. The key factors are the limited processing power and imperfection of IoT's applications and the lengthy periods with which protection applications areupgraded.The protection of current IoT networks is, thus, one of the key guidelines for improving in-service security technologies at present. Software bugs and their early fixes generate a threat to irregular behavior on IoT computers. It suggests that modern and more efficient approaches [3] for detecting and counter measuring such attacks are desperately needed.

The use of machine learning techniques is one of the popular modern approaches for identifying suspicious events and attacks within computer networks.

The paper's contribution is the following

- It provides an overview of the issue regarding the combined use of machine learning techniques and concurrent data processing to detect attacks.
- A system architecture that enables anomalous events to be observed on IoT devices depending on the solution suggested. Compared with the classificatory mixture system, a weighted composition of classifications at the last stage of the curriculum allows the possibility of simultaneous training for the simple classifiers and increased the accuracy performance.
- The analysis discussed in this paper aims to measure the accuracy rate from the perspective of binary attack detection on IoT Networks based on the estimate of a particular class of potential aggressions. Conclusions are made on the possibilities of the method suggested for further development based on the interpretation of data obtained.

2. RELATED WORK

In the research community, the use of machine learning and parallel data technology is widespread. Despite this, a limited amount of work is carried out in the framework of IoT [1] to address the difficulties of concurrent data processing frameworks, e.g. DryadLINQ, Hadoop, Spark and MPI.

DryadLINQplatform is Microsoft platform that offers parallel and distributed software creation tools [2]. This application uses specific expressions that convert a dataset. The execution of the programmed is structured as an acyclic graph, with vertices as processes and borders as interprocess channels.

Hadoop introduces map Minimize principle. Data analysis is done in two steps accordingly. The first stage (map) involves the isolation and propagation of the input data between the device nodes. The second stage (reduce) requires the simultaneous collection of data and aggregation on these nodes.

Spark is a computational platform for resilient and concurrent distributed data sets (RDDs). Data RDD is grouped into blocks so that we can control them on many devices independently. One of the principal properties of the RDD is that if one computer fails, can recover the independent partitions. It provides the cluster based on Spark of fault tolerance. In comparison to Hadoop, Spark gives the ability to store memory results and slow computational methods. It helps you to gain the speed at which iterative algorithms are performed, particularly in machine learning algorithms.

MPI (Message Forwarding Interface) is another method for creating parallel structures. The MPI's basic unit is the message. Two methods are used in messages for the contact between processes: transfer and receipt. Because of the low levels of parallel processes, a high data processing speed is given. The popularity of this technology when designing supercomputers is largely decided.

In several academic papers, the challenge of designing parallel methods of processing data is solved and extended to different fields. To detect the malicious malware in mobile application network, Shamily, Bauckhage and Alpcan, in particular, are suggesting implementations of Distributed Support Vector Machine (SVM). To this reason, the distributed algorithm has been developed, which underlies the SVM setup to resolve the quadratic binary classification problem. Scherbakov et al. recommend a web-server Apache log review framework. The architecture [9] of the device consists of three stages, including interpretation, business logic and data layers. Kim and Yu suggest introducing CEP framework, which is designed for bus traffic control and integrates Earthquake, Esper and Hadoop. Similar equipment is also used in the study of medical results.

Many mechanisms cross the topics of machine learning and concurrent data processing. TensorFlow, DistBelief, MXNet, and Piccolo are thus oriented not to an overall distributed computing paradigm applied on the Hadoop and Spark systems, but to solving machine learning issues under distributed computing conditions. One common feature of these systems store and control the mutual state using parameter servers. Shared state is expected for multiple machines to scale model training. Besides, in a single dataflow diagram, TensorFlow describes all machine learning tasks with parameters. This architecture varies from parallel systems in two ways for general purposes to boost machine learning performance. Next, the model maintains many runs on overlapping sub diagrams of the overall diagram. Application for IoT authentication, malware identification and access management regarded the numerous methods of supervised, unregulated learning.

Work has been provided on deep learning approaches to detect cyber threats [4] [5] [6] [7]. The proposed method consists of several steps:

(1) The main analysis part is applied;
(2) The neural network is pretrained using a restricted Boltzmann machine;
(3) Deep neural network training;
(4) Output signal is generated based on a Softmax regression.

Approaches are also systematically studied to develop parallel data processing structures. However, the use of machine learning techniques is improperly evaluated in these applications to ensure the security of IoT users. This distance is being eliminated by the solution suggested in this article.

3. METHODOLOGY &FRAMEWORK

DATASET DESCRIPTION:

In the dataset for experiments «Detection of IoT Botnet Attacks». There are records in this data collection reflecting network stream vectors amongst nine commercial IoT devices. Two botnets have created anomalous network traffic: Mirai and BASHLITE.

The dataset comprises 7009270 documents separated into a set of classes: a range of attack classes and a benevolent class of traffic. The following classes comprise a selection of attack classes: ack_Mirai, scan_Mirai, syn_Mirai, udp_Mirai, udpplain_Mirai, combo_BASHLITE, junk_BASHLITE, scan_BASHLITE, tcp_BASHLITE, and udp_BASHLITE. It is found that these attacks to be the most common for IoT and most traditional to validate the approach to attack detection based on computer and Big Data [8]approaches.The format of the display of documents is CSV: 115 fields (network stream features) are separated by a comma for each record.

DATA TRAINING:

In this paper, the method for creating training samples to improve the precision of the attack classification is developed. Duplicate data is omitted in the first place. And the data, which are weakly correlated, are retrieved. The Pearson correlation coefficient determines the degree that things are identical.

DATA LOADING AND SAMPLE FORMATION:

The records are put in 11 CSV files for each computer. Per file is equal to one of eleven grades. The Python programming language and Spark Data Frame API have been chosen to allow for efficiently storing and processing of CSV files.

The Spark Session entry point has been used to build the data frame object. The object DataFrameReader is used to load data from external storage into the data structure (e.g. object-relative databases, file systems, key-value stores). API helps you to load data from files with formats CSV, JSON, PARQUET, TEXT, and JDBC.

MODEL TRAINING, TESTING AND EVALUATION:

Experiments investigated a limited number of common classificatorytypes: DecisionTree (DT), Random Forest, DNN, support vector machine (SVM) and Extreme Machine Learning (EML).The MLlib library implements these models with their learning algorithms. There are a vast variety of data analytical functions obtained in this collection, using machine learning and mathematical techniques. These features are designed for distributed mode execution.

4. EXPERIMENTAL RESULTS

In this paper, performance for various classical algorithms such as SVM, Random Forest and Naive Bayes etc. to detect attacks on the network using IDS datasets such KDD, NSL is evaluated. However, dynamic attacks cannot be predicted using these classical algorithms if attacker introduces a new attack with changing parameter. Therefore, algorithms need to train in advance to overcome this problem. In this paper, the author has evaluatedthe performance of Deep Neural Network (DNN) algorithm with dynamic attack signatures and detection accuracy of DNN,as shown in Figure 1. It is compared with other classical algorithms.

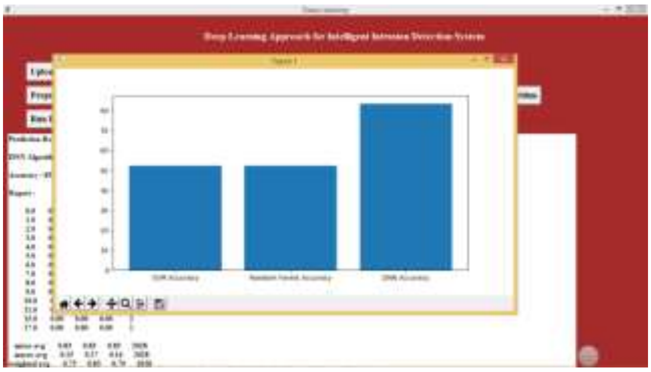
Deep Learning Approach for Intelligent Intrusion Detection System
SVM Accuracy
Random Forest Accuracy
DNN Accuracy

Figure 1: DNN with other Classical Algorithms

The graph x-axis displays the algorithm name and y-axis, and the DNN is a more precise technology. Parallel processing techniques for effective online load forecasting, which takes less time compared to other classification algorithms.

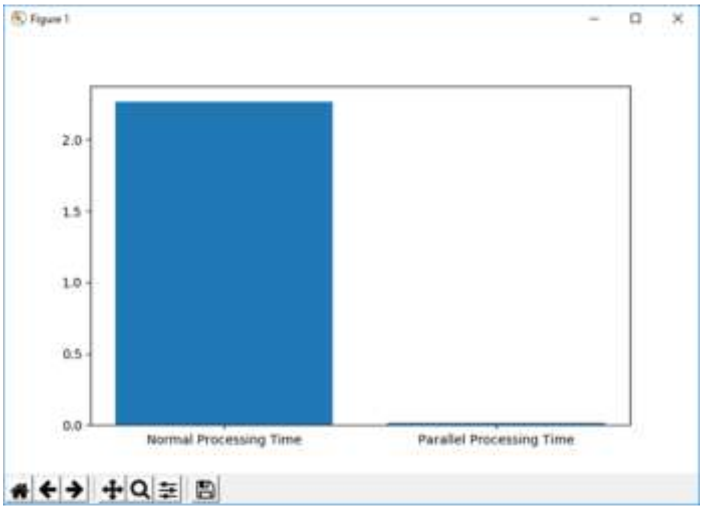

Figure 2: Time required to Normal vs Parallel Processing

The graph in Figure 2 shows that parallel processing takes less time than normal processing.

5. CONCLUSION

A new approach has been suggested in the paper to classify attacks on IoT devices focused on machine learning and data analysis. The layout of simple classifiers for attack detection in IoT networks was calculated on the basis of the study of cutting-edge implementations of machine learning techniques and parallel data processing for the solution of computer security problems. It involves SVM, RF, DNN and EML Based on simultaneous data processing; a classification issue statement was developed in which the fundamental efficacy measure was the accuracy-to-time ratio for preparation and research.

An experimental test of the proposed solution found that the precision and speed of attack detection in the IoT Network has substantially increased. The sensitivity is approximately 100%, and during the identification, the speed increases due to the number of parallel threads. The use of methods for integrating simple classifiers will improve the accuracy of the basic classifiers shown.

REFERENCE

[1] D. Evans, "The Internet of Things How the Next Evolution of the Internet Is Changing Everything," CISCO white paper, 2011.

[2] Zh.J. Shi and H. Yan, "Software Implementations of Elliptic Curve Cryptography," International Journal of Network Security, vol. 7, no. 1, pp. 141–150, July2008.

[3] M. Yassineand A. Ezzati, "Towards an Efficient Datagram Transport Layer Security for Constrained Applications in Internet of Things," International Review on Computers and Software, vol. 11, no. 7, pp. 611- 621, 2016, doi:10.15866/irecos.v11i7.9438.

[4] G. Apruzzese, M. Colajanni, L. Ferretti, and A. Guido, "On the Effectiveness of Machine and Deep Learning for Cyber Security", in Proc.ofthe 10th International Conference on Cyber Conflict (CyCon), pp. 371– 390,2018,doi:10.23919/CYCON.2018.8405026.

[5] Th. Nguyen and V.J. Reddi, "Deep Reinforcement Learning for Cyber Security", CoRR, http://arxiv.org/abs/1906.05799,2019.

[6] D.S. Berman, A.L. Buczak, J.S. Chavis, and Ch.L. Corbett, "A Survey of Deep Learning Methods for Cyber Security", Information, vol. 10, no. 4, 122, https://www.mdpi.com/2078-2489/10/4/122, 2019, doi: 10.3390/info10040122.

[7] M. Usman, M.A. Jan, X. He, and J. Chen, "A Survey on Representation Learning Efforts in Cyber security Domain", ACM Comput. Surv. vol. 52, no. 6, Article 111, 28 pages, October 2019, doi: 10.1145/3331174.