An Efficient Machine Learning Techniques Based on IoT for Effective Load Forecasting

K. Rameshwaraiah$^{1}$ Sankarpu Sateesh Kumar$^2$ K. Srinivasa Babu$^3$ T. Madhu$^4$

$^1$Professor Computer Science and Engineering
$^2,4$Assistant Professor Computer Science and Engineering
$^1,3,4$Nalla Narasimha Reddy Educational Society’s Group of Institutions
$^2$RGUKT – Srikakuram, A.P

Abstract — Internet of Things (IoT) networks (IoT) are computer networks which have an acute IT protection problem and in particular a computer attack detection problem. In order to solve this issue, the paper recommends the combination of machine learning approaches and concurrent data processing. The framework is developed and a new approach to the combination of the main classifiers intended for attacks on IoT networks. In which the accuracy ratio to the training time is the integral measure of efficacy, the problem classification statement is developed. We recommend the use of the data processing and multi-threaded mode provided by Spark to accelerate the speed of training and testing. In addition, a technique is suggested for preprocessing data set, which results in a large reduction in the sample length. An experimental examination of the proposed approach reveals that the precision of IoT networks’ attack detection is 100 percent and the processing speed of data sets increases proportionally to the number of parallel threads.

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

I. Introduction

The concept of the Internet of Things (IoT) is now being used increasingly in many areas of computer and technology engineering. Hospitals, residential administration, automation, traffic control, transport by rail and air, energy usage, development, etc. There is a clear 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 (WiFi, sensors, cell networks, and other networks) into a common global computer network. They are therefore linked to different types of networks and networks. This ensures the interchange of products and conventional computing equipment. Other characteristics of IoT networks include the high heterogeneity of its components, geographic distribution and a 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 is upgraded. 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 behaviour on IoT computers. This suggests that modern and more efficient approaches for detecting and countermeasuring 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 combined use of machine learning techniques and concurrent data processing to detect attacks.
- System architecture that enables anomalous events to be observed on IOT devices depending on the
solution suggested. Compared with the classifier 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. A new method of generating a training sample for machine learning algorithms, focused on the usage of parallel computing, is presented. Compared to [8], the implementation of this method is also performed for the local mode.

- 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 certain class of potential aggressions. Based on the analysis of results obtained, conclusions are drawn regarding the possibilities of the approach proposed for further development.

II. 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 to address the difficulties of concurrent data processing frameworks, e.g. DryadLINQ, Hadoop, Spark and MPI.

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

The Map Minimize principle is introduced by Hadoop. 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 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, the independent partitions can be recovered. This 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. This 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. In order to detect the malicious malware in a mobile application network, Shamily, Bauckhage and Alpcan in particular are suggesting implementations of the 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 of the device consists of three stages, including interpretation, business logic and data layers. Kim and Yu suggest introducing the 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.

There are many mechanisms that cross the topics of machine learning and concurrent data processing. TensorFlow, DistBelief, MXNet, and Piccolo [24, 25] 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 is that they store and
control the mutual state using parameter servers. Shared state is expected for multiple machines to scale model training. In addition, 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 in order to boost machine learning performance. Next, the model maintains many runs on overlapping subdiagrams 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. The proposed method consists of several steps: (1) the main analysis part is applied; (2) the neural network is pretrained using a restricted Boltzman machine; (3) the deep neural network training; (4) the 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.

III. METHODOLOGY & FRAMEWORK

Dataset Description:

We have used 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. We found these attacks to be the most common for IoT and the most traditional to validate the approach to attack detection based on computer and Big Data approaches.

The format of the display of documents is CSV: 115 fields (network stream features) are separated by a comma for each record. Each field is not empty and has a valued form that allows the use of this data set in classification and clustering.

Data Training:

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

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 DataFrame API have been chosen to allow for efficient storing and processing of csv files.

The SparkSession entry point has been used to build the dataframe 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). The API helps you to load data from files with formats CSV, JSON, PARQUET, TEXT, JDBC.

Model Training, Testing and Evaluation:

The experiments investigated a limited number of common classificator types: decision tree (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.

IV. EXPERIMENTAL RESULTS

In this paper author is evaluating performance of various classical algorithms such as SVM, Random Forest and Naive Bayes etc. to detect attacks on network using IDS datasets such KDD, NSL but this classical algorithms unable to predict dynamic (if attacker introduce new attacks with changes in attack parameter) attacks and need to be trained in advance to detect such attacks to overcome from this problem author has evaluate performance of Deep Neural Network (DNN) algorithm with dynamic attack signatures and detection accuracy of DNN shown to be better compare to all classical algorithms.

In above graph x-axis represents algorithm name and y-axis represents accuracy and DNN is more accurate technique.

We use parallel processing techniques for effective online load forecasting, which takes less time compared to regular loading The graph above shows that parallel processing takes less time than normal processing.

V. CONCLUSION

A new approach to detect attacks on IoT devices based on machine learning and parallel data processing was proposed in the paper. 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. This 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 solution proposed found that the precision and speed of attack detection in the IoT Network was substantially increased. The sensitivity is approximately 100% and during the identification the velocity 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.

REFERENCES

[1] D. Evans, “The Internet of Things How the Next Evolution of the Internet IsChanging 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, July 2008.
[3] M. Yassine and 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. of the 10th International Conference on Cyber Conflict (CyCon), pp. 371–390, 2018, doi:10.23919/CYCON.2018.8405026.
[5] Th.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.
[8] I. Kotenko, I. Saenko, A. Kushnerovich, and A. Branitskiy, “Attack detection in IoT critical infrastructures: a machine learning and big data processing approach”, in Proc. of the 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pp. 340-347, 2019, doi:10.1109/EMPDP.2019.8671571.
[9] Kotenko I.V., Saenko I.B., Kushnerevich A.G., “Architecture of the Parallel Big Data Processing System for Security Monitoring of Inter-net of Things Networks”, SPIIRAS Proceedings. Issue 4(59). pp.5-30, 2018, doi:10.15622/sp.59.1.