International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Evaluation of Hybrid Deep Learning Anomaly Detection Approach for Attacks Over Network

[1] Sameeksha Gupta, [2]Pankaj Richhariya

Department of Computer Science
Bhopal Institute of Technology & Science, Bhopal

ABSTRACT-

The number and severity of network attacks that breach users sensitive and important data are rising, particularly in light of the increasing integration of the Internet into our everyday lives. The majority of conventional defense strategies have been rendered less effective due to the sheer number and complexity of these entry attempts. At the same time, the cyber security community became interested in doing research to improve and automate intrusion detections due to the impressive results of machine learning techniques, particularly deep learning, in computer vision. However, it is difficult to adequately oversee the training of an intrusion detector due to the high cost of data labeling and the limited of anomalous data. As a result, another crucial aspect is intrusion detection based on unsupervised anomaly detection. In this study, we present a three-stage framework for network intrusion attack detection based on deep learning anomaly detection. The framework integrates techniques for supervised learning (CNN), semi-supervised learning (GANomaly), and unsupervised learning (K-means clustering). The proposed algorithm worked on the parameters like-accuracy, precision, time and shows better performance.

Index Terms—*Anomaly Detection, Intrusion Detection, Deep Learning, Unsupervised Learning, Neural Networks.*

1. Introduction

The process and method of identifying a deviation or an unusual fact in a given dataset is called ***Anomaly Detection***. Sometimes anomalies hold appreciated information about irregular features of the systems [1]. When an arrangement's behavior deviates from how it was previously operating, it may indicate unanticipated behaviors that were previously unnoticed; occasionally, these behaviors may be harmful, but other times, they may not. An outlier is another synonym for anomaly. While working on data mining, many academics typically concentrate on other techniques like clustering and classification. However, the field of assessment saw a revolution when academics and researchers began to identify peculiar objects that might be used to address real-world problems with damage detection, fraud and scam detection, anomalous medical status identification, and intrusion detection.

Anomalies can be divided into three categories: point, contextual, and collective anomaly. A specific event is classified as a point anomaly if it can be quantified as abnormal with regard to any of its features. For example, if a certain data instance is irregular. The anomaly appears in a certain area or at a specific time. Collective anomalies can be defined as a grouping of related data instances that are abnormal with respect to the entire dataset, but not with respect to distinct criteria. When anomaly detection is based on whimsy, it is most accurate; anomaly detection can be used as reminder support in the future [2].

The Hybrid Anomaly Detection (EEHAD) approach needs to cognize the ordinary conduct of a client’s gadget is proficient to separate amongst typical and anomalous application likely malevolent activities.

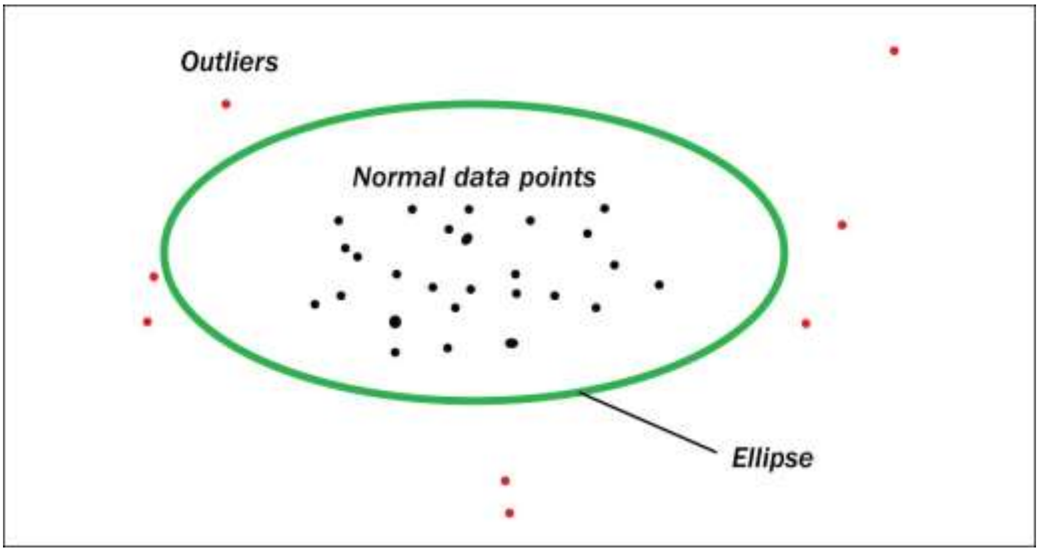

Figure 1-Anomaly Points.

1.1 Detecting Anomaly in IoT

Since, machine learning techniques are accurate and efficient at analyzing vast volumes of data, they have found widespread application in the Internet of Things for anomaly identification. 5Many machine learning approaches, including supervised ones like support vector machines (SVMs) and decision trees as well as unsupervised ones like clustering algorithms and autoencoders, have been used for anomaly detection in the Internet of Things [2–4]. Numerous data kinds, such as sensor data, network traffic data, and financial transaction data, may be processed using these methods.

It is possible to detect anomalous actions in the IoT environment, identify malicious data, and stop network assaults by using several anomaly detection algorithms to find patterns in IoT data streams. These days, a number of methods—including statistical-based [2], proximity-based [3], machine learning (ML) based [4], and deep learning (DL)-based approaches have become widely employed for IoT anomaly identification. The massive volume of data produced by IoT devices is a problem for machine learning-based anomaly detection in the Internet of Things. Machine learning algorithms may find it challenging to appropriately identify abnormalities in this data due to its potential for noise, high dimensionality, and heterogeneity.

Furthermore, because of the IoT's dynamic nature, data linkages and patterns may alter over time, necessitating frequent updates or retraining of the machine learning model. A further obstacle to anomaly identification in the Internet of Things is the scarcity of labeled training data. Getting a big enough dataset with both typical and abnormal samples is often difficult to do in order to properly train a machine learning model. This can be particularly troublesome in cases when anomalous occurrences are uncommon or unpredictable. Machine learning algorithms have demonstrated potential for anomaly identification in the Internet of Things, despite these obstacles. However, a specific industrial application is frequently the focus of existing literature articles on anomaly detection in the Internet of Things utilizing machine learning algorithms.

2 Review of Literature

The author of this study introduces lightweight intrusion detection systems (IDS) that are supplemented with a strong MLP neural network to identify anomalous activities in Android mobile devices. The IDS gathers several NetFlow characteristics and keeps an eye on the mobile device's network activity continually. After gathering the data flows, an artificial neural network (ANN) decides whether or not there has been an invasion. The outcomes of the experiment show that IDS is capable of successfully identifying an abnormality in the Android operating system [1].

The author of this study described a data mining-based intrusion detection system for smartphones, and then used real-time user data to assess the technology's effectiveness. The novel Intrusion Detection Architecture for Mobile Networks (IDAMN) approach is provided by the author. The efficacy of this approach in identifying intrusions was demonstrated. Next, the effectiveness of two classifiers—Naïve-Bayes and SVM—was examined. It is quite successful for both classifiers to identify intrusions [11].

In this paper author presented and demonstrated utility of the interruption discovery idea, the location segment is starting to get more consideration. There are numerous ways to deal with discover interruptions under location segment. Since volume of information managing system is so substantial, this accumulation focuses vigorously on the utilization of information mining in the region of interruption recognition. Grouping is one of the compelling strategies under information mining that can be utilized for interruption discovery. Two basic information portrayals for arrangement method are IF-THEN principles and Decision tree [12].

An unsupervised machine learning technique called clustering seeks to organize input datasets into groups called clusters, each of which contains items that are similar to each other and different from those in other clusters. In IoT applications, clustering-based techniques are frequently employed for anomaly identification. The efficacy and accuracy of combining clustering approaches to identify outliers in an input dataset have been demonstrated by several studies. For example, Alguliyev et al. [21] developed a hybrid model that combines the K-means and particle swarm optimization (PSO) methods to accurately find abnormalities in huge data.

In order to decrease intra-cluster distances and maximize inter-cluster distances, a unique weighted clustering algorithm is presented in this study [21]. Many local minimum points and preset cluster centers are eliminated by the suggested strategy. The performance and clustering accuracy of the suggested technique outperform those of the conventional K-means algorithm, as demonstrated by the experimental findings, which are based on a single dataset. Clustering techniques usually need the user to provide a set of parameters.

Rahman et al. [22] describe a unique density based clustering approach for outlier identification that is parameter independent and can cluster data of different forms, since K-means clustering presupposes clusters of spherical shapes. There are two steps to the proposed parameter-independent density-based clustering (PIDC) method. Using unique nearest neighbor principles, the first step finds outliers in the datasets and eliminates them. Following that, the records move on to the density-based clustering step two. Six datasets are used to test this study, and the suggested algorithm's performance is compared to five popular clustering techniques using comparative analysis. On high-dimensional datasets containing outliers, the suggested technique works well, but its computing complexity is significant.

Real-time data necessitates quick outlier detection when it comes to anomaly detection in Internet of Things applications. Rapid data delivery necessitates quick calculation with little memory use [23]. To solve these issues, Bah et al. [23] introduce a unique hybrid model called Micro-Cluster with Minimal Probing (MCMP), which combines Thresh_LEAP with Micro-Cluster Outlier Detection (MCOD). The goal of this study is to reliably detect distance-based outliers while minimizing computing performance and memory usage. In order to reduce computations based on distance, micro-clusters hold nearby data points [23], which reduces processing time and memory use.

Yang et al. [24] provide an enhanced Self-Organizing Feature Map (SOFM) technique to cluster the dataset, and then propose an outlier identification algorithm called the *Neighbor Entropy Local Outlier Factor (NELOF)*. The enhanced SOFM algorithm modifies the neurons dynamically and avoids the random selection of neurons by utilizing the Canopy algorithm. The number of dead neurons is dramatically reduced btry the enhanced SOFM method, which outperforms the original SOFM algorithm. Moreover, in order to lessen the impact of outliers in K distance neighborhoods, this study substitutes relative K-distance neighborhoods for K-distance neighborhoods [24].

The suggested NELOF method performed better than the LOF technique in terms of execution time and accuracy when evaluated on seven different datasets; nevertheless, the usefulness of the suggested approach on high-dimensional datasets is not investigated in this study. A Fog-empowered anomaly detection approach employing hyper ellipsoidal clustering (HyCARCE) is presented by Lyu et al. [25] because anomaly detection is essential for time-sensitive IoT applications that include high-dimensional data. In order to enable cloud offloading, fog computing is proposed as a substitute for cloud computing [25].

By reducing the computational overhead at the sensors and the cloud, the Fog architecture improves detection responsiveness and time delay by having the end nodes or sensors send data directly to the Fog nodes for clustering and anomaly detection. HyCARCE capacity to automatically choose the number of clusters and its adaptability to various data distributions, such as linear or hyper spherical, are two benefits of utilizing it for data clustering. It is demonstrated that the suggested technique can reliably and quickly identify outliers and anomalous clusters in the dataset by applying it to two synthetic and two real-world datasets. Because of the information exchange in the Fog architecture, privacy and security concerns are not addressed in this study.

A deep learning-based intrusion detection system for the Internet of Things was proposed by Almiani et al. [26]. The traffic analysis engine and the categorization engine are the two main parts of their model. The pre-processing of the traffic data, including feature reduction, normalization, and symbolic-to-numeric transformation, is done by the traffic analysis engine. The classification engine then receives the processed input and uses two deep recurrent neural networks (RNN) to reply quickly in a real-time setting. The two RNNs function as two attack detection filters. The second RNN detection layer will receive the traffic data that the first RNN layer assessed as normal in order to determine whether or not it is an abnormality. The two RNN layers are trained using the same dataset. The training sets of the two RNNs differ mainly in that the first RNN training set includes both abnormal and normal traffic data, whereas the second RNN training set exclusively includes normal traffic data.

To identify abnormalities in IoT data, Amrit Poudel [23] introduced the Vector Convolutional Deep Learning (VCDL) model, which makes use of a recently developed distributed intelligence technique known as "*Fog Computing*." Three tiers of components make up the suggested model. Distributed Internet of Things devices make up the initial layer. The second layer is called the fog layer; it consists of many work fog nodes that are networked to IoT devices and that train every VCDL model in a distributed fashion. In the fog layer, the worker nodes will receive and share the optimal set of parameters from the master fog node. As a result, the matching worker node will get the traffic data and classify it as normal or attack. The third layer of the suggested structure, the cloud layer, will receive the categorization result. The data from the whole fog layer is verified using the cloud layer. The experimental findings show that, in comparison to the centralized detection approach, the suggested distributed VCDL architecture can identify anomalous traffic data with more accuracy and in a shorter amount of time.

3. Problem Formulation

1. The previous given solution such as pattern detection is based on the anomaly signature identification. It needs to maintain an updated database which will be further building for intrusion behavior detection.

2. A behavior based technique is also another technique which is limited to analyze the application behavior, its data accessing methodology, accessing of multimedia and textual personal data. But still unable to access the memory level occupancy of application. The existing technique need to get the proper training before the evaluation.
3. A proper evaluation under the given Anti-malware technique unable to detect multiple instances.
4. Previously given technique does not involve any energy depletion analysis module, thus an analysis of energy consumption by the application is not considered.
5. Previously given approach not involved their efforts in battery optimization and battery monitoring application.
6. A limited and single criteria based approach solution is given by previous literature discussed author.

As per the above discussion point, it is enabling to understand how the different factor affects smart phone efficiency. The previously given technique nowhere help in battery optimization. Thus the further work is provided in this direction.

4. Proposed Method

The main goal of effective IDS is to provide high rates of attack detection with very small rates of false alarms. An approach which is working towards the energy optimization and battery monitoring is utilized in proposed technique HAD. The memory utilization, CPU observation are performed to help in Intrusion application detection monitoring. The key components of these systems are:

- **Information Source**: Data utilized by the IDS.
- **Analysis Engine**: Process by which the intrusion Detection is made. The overall analysis of battery usage detection.
- **Response**: Action taken when an intrusion is detected.

The principal module is the analysis engine. The analysis engine applies three types of techniques for analyzing and detecting a security threat. These methods are the battery monitoring, energy depletion frequency, the signature based technique, the anomaly detection, and the protocol anomaly detection. In our approach, the second method is adopted, which usually depends on an Hybrid Anomaly Detection (HAD) Approach. The Hybrid Anomaly Detection (HAD) Approach is used for detecting unknown threats. For this purpose, it is prior trained with the aim of normal and malicious traces. In the context of this thesis a Rule Mining Technique (RMT) is proposed due to its light-weight operation and its efficiency.

4.1 Algorithm

The methodology has major five components:

1. Networking Monitoring Module
2. Flows Extraction Module
3. Feature Extraction Module
4. Analysis Engine Module Using battery optimization technique
5. Alert Module

In the given above solution proposed is based on battery optimization which help in complete mobile app activity analysis. The proposed HAD approach given solution help in optimal analysis of application usage, which help in intrusion application detection. The given pseudo code help in understanding of the work flow execution.

5. Result Analysis

Finally, the result has been finding out by using the parameters, on the basis of the existing algorithm and the proposed algorithm.

In the results figure shows various parameters and their outputs on the basis of existing and proposed algorithm.

The experimental results indicate that our IDS can detect anomalies of the Android system with relative accuracy and detection rate.

- **Total Netflows:**

Netflow is a feature on Cisco network devices that allows you to Collect & Analyze network traffic that goes through a Router/Switch. An IP Flow is made up of a Set of 5 attributes and can have up to 7 totals.

- **Malicious Netflows:**

We presented lightweight IDS which are enhanced with a powerful MLP neural network for detecting malicious behaviors of the Android mobile devices.

- **Accuracy:**

At the point when an interruption is shown accurately, we have a "Genuine Positive" (TP) truth. At the point when a no interruption is demonstrated and this attestation is right, we have a "Genuine Negative" (TN). At the point when the IDS demonstrate an interruption and this declaration isn't right a "False Positive" (FP) caution is activated. Finally, when a non-interruption is demonstrated and an interruption is in reality in advance, we have a "False Negative" (FN) occurrence. FN is the most pessimistic scenario circumstance of each recognition system since it causes a false alert. Given these terms, we assessed our IDS utilizing the exactness esteem and the identification rate. Precision (ACC) is characterized in the accompanying condition as the quantity of interruptions over the aggregate number of occasions.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}$$

- **Detection Rate:**

On the other hand, the detection rate (DR) is the probability of an alarm given of all the actual intrusions.

$$\mathrm{DR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

5.1 Statistical Analysis

According to the aforementioned definitions, the results of our experiments are summarized in Table 6.1. The experimental results indicate that our IDS can detect anomalies of the Android system with relative accuracy and detection rate to 78.23% and 87.27% respectively.

Table 1-Result Analysis of Existing Algorithm with Proposed Algorithm

Parameters	Existing Algorithm	Proposed Algorithm
Computational Time	550ms	443ms
No. of Detection	5	
Accuracy	78.23%	87.27%
Precision	63%	75%
Recall	57%	89%
F-Measure	64%	81%

In the table given above shows the difference analysis between the existing technique and proposed battery based approach. It helps in understanding the efficiency of proposed HAD technique.

5.2 Graphical Comparison Analysis

In this section an analysis of result is presented, the section gives an understanding of statically graphical analysis.

Graph for the computational Analysis

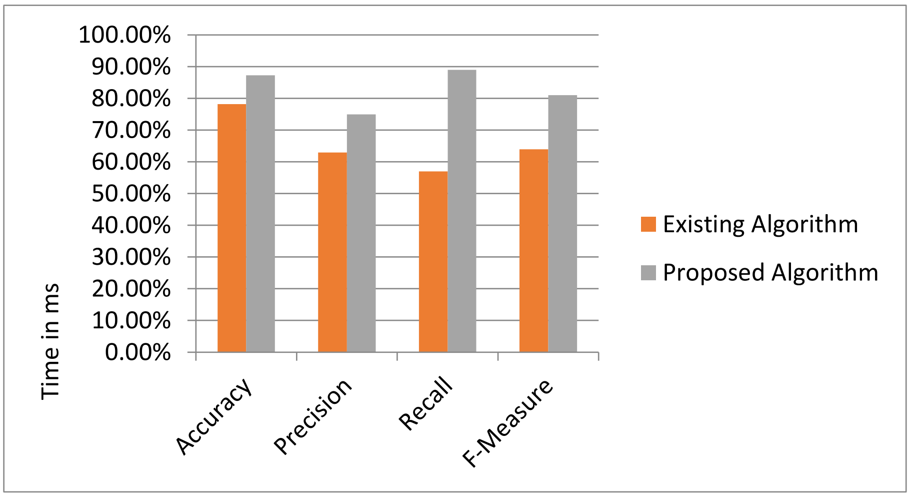

Figure 2-Comparison between the existing intrusion application analysis and proposed solution which is battery based.

As presented in the figure 2 the energy consumption by various smart phones has been compared by time in milliseconds.

Graph for the computational Analysis

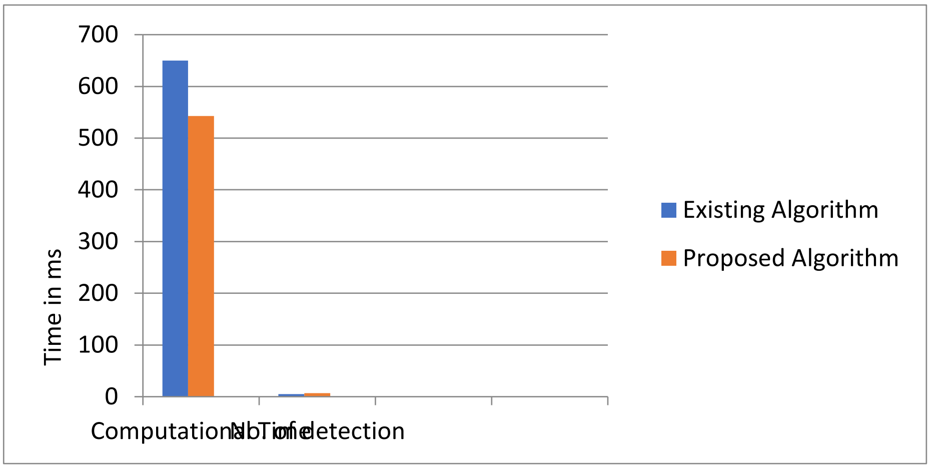

Figure 3-Comparison between the existing intrusion application analysis and proposed solution which is battery based.

As presented in the figure 3 the energy consumption by various smart phones has been compared by time in milliseconds.

6. Conclusion

Modern mobile phone utilization is logically progressing in the current subjective apps and administrations. Some apps, such email checking, browsing the Internet, participating in casual groups, and so on, have advanced to the point where they are now part of the capabilities of mobile phones. More digital attacks, such as the creation of adaptable malware for various uses, have been drawn in by the enhanced computational and capacity capabilities of modern smartphones. Hackers and other cybercriminals are drawn to the sophisticated features and widespread use of contemporary mobile devices. In this work, we introduced the use of Rule Mining Technique (RMT) in Hybrid Anomaly Detection (HAD) Approach to identify anomalies related to harmful behaviors of Android mobile devices based on energy consumption time, as based on IDS. The results of the experiment show that our IDS is capable of efficiently identifying anomalies in the Android operating system and improving the algorithm that handles the behavior structure analysis and battery analysis of mobile applications on Android devices. The trial findings show that our IDS is capable of successfully identifying an abnormality in the Android operating system. In particular, the suggested system's accuracy and detection rate are 87.27% and 78.23%, respectively.

7. Future Work

The correctness of the suggested IDS was evaluated using a monitoring approach to prior work; although the suggested algorithm is secure, it is still possible to improve it further in the ways listed below:

1. Future research attempts to increase the IDS's accuracy and detection rate by accounting for more crucial traffic characteristics, such as user touch patterns and behaviors.
2. Reducing the energy usage of mobile phones will be a top goal in the future.

REFERENCE

[1] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, and K. Mankodiya, "Towards Fog-Driven IoT eHealth: Promises and Challenges of IoT in Medicine and Healthcare," Future Generation Computer Systems, vol. 78, 2018, pp. 659-676.

[2] Aubrey-Derrick Schmidt, Frank Peters, Florian Lamour, Christian Scheel, Seyit Ahmet amtepe, S¸ahin Albayrak, “Monitoring Smartphones for Anomaly Detection” http://www.springerlink.com DOI:10.1007/s11036-008-0113-x.

[3] Vaibhav Rastogi, Yan Chen, and Xuxian Jiang “Evaluating Android Anti-malware against Transformation Attacks” (March 2013).

[4] Jess Stubenbord, “Battery Information Display in Mobile Devices”.

[5] Apeksha Vartak, Darshika Pawaskar, Suraj Pangam, Tejal Mhatre, Prof. Suresh Mestry, “HYBRID INTRUSION DETECTION USING SIGNATURE AND ANOMALY BASED SYSTEMS”, International Journal of Innovative Research in Science and Engineering Vol. No.2, Issue 03, March 2016. www.ijirse.com.

[6] Muhamed Halilovic, Abdulhamit Subasi, “Intrusion Detection on Smartphones”.

[7] JABEZ J, Dr.B.MUTHUKUMAR, “Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach”, Procedia Computer Science 48 (2015) 338 – 346.

[8] P. Garcı´a-Teodoro, J. Dı´az-Verdejo, G. Macia´-Ferna´ndez, E. Va´zquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges”, computers & security (2009) 18 – 28, www.elsevier.com/locate/cose.

[9] Ming syan chen & Jiawei han (Senior Member IEEE) and Philip S. Yu, fellow IEEE, “Data Mining: An Overview from a database Perspective”, IEEE Transaction on knowledge and data engineering, Vol.8, No.6, December 1996.

[10] Bing Liu, Wynne Hsu, Yiming Ma, “Integrating Classification and Association Rule Mining”, From: KDD-98 Proceedings 1998, AAAI www.aaai.org.

[11] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” IBM Journal of Research and Development, vol. 3, no. 3, 1959, pp. 210–229.

[12] M. AlDarwish, “Machine Learning,” ML. [Online]. Available: http://www.contrib.andrew.cmu.edu/~mndarwis/ML.htm.

[13] Guillermo Suarez-Tangil, Juan E. Tapiador, Flavio Lombardi, Roberto Di Pietro, “ALTERDROID: Differential Fault Analysis of Obfuscated Smartphone Malware” SUBMITTED TO IEEE TRANSACTIONS ON MOBILE COMPUTING.

[14] Ali Raza, “A battery and Network Usage Model for Smartphones”, (1st June 2012) Faculty of Engineering and Science University of Agder. Androidology: Architecture overview http://developer.android.com/videos/index.html#v=QBGfUs9mQYY.

[15] Turkka Salmi, “Energy efficient use of the Smartphone’s”, (2017) Oulu University of Applied Sciences Degree programme in Business Information Technology.

[16] Cristina L. Abad, Yi Lu, Roy H. Campbell, “DARE: Adaptive Data Replication for Efficient Cluster Scheduling”, 2011 IEEE International Conference on Cluster Computing.

[17] Earlence Fernandes, Bruno Crispo, Senior Member, IEEE, and Mauro Conti, Member, IEEE “FM 99.9, Radio Virus: Exploiting FM Radio Broadcasts for Malware Deployment”, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 8, NO. 6, JUNE 2013.

[18] Haoyu Li, Di Ma, Nitesh Saxena, Babins Shrestha and Yan Zhu, “Tap-Wave-Rub: Lightweight Malware Prevention for Smartphones using Intuitive Human Gestures”, WiSec’13, April 17-19, 2013, Budapest, Hungary. Copyright 2013 ACM 978-1-4503-1998-0/13/04.

[19] Karthikeyan .K.R and A. Indra, “Intrusion Detection Tools and Techniques A Survey”, International Journal of Computer Theory and Engineering, Vol.2, No.6, December, (2010) 1793-8201.

[20] Dimitrios Damopoulos, Sofia A. Menesidou, Georgios Kambourakis, Maria Papadaki, Nathan Clarke2 and Stefanos Gritzalis, “Evaluation of Anomaly-Based IDS for Mobile Devices Using Machine Learning Classifiers”, SECURITY AND COMMUNICATION NETWORKS Security Comm. Networks 2011; 00:1–9 DOI: 10.1002/sec.

[21] Rasim M. Alguliyev, Ramiz M. Aliguliyev, and F. J. Abdullayeva, "PSO+K-Means Algorithm for Anomaly Detection in Big Data," Statistics, Optimization & Information Computing, vol. 7, no. 2, 2019, pp. 348.

[22] M. A. Rahman, K. L. Ang, and K. P. Seng, "Unique Neighborhood Set Parameter Independent Density-Based Clustering with Outlier Detection," IEEE Access, vol. 6, 2018, pp. 44707-44717.

[23] M. J. Bah, H. Wang, M. Hammad, F. Zeshan, and H. Aljuaid, "An Effective Minimal Probing Approach with Micro-Cluster for DistanceBased Outlier Detection in Data Streams," IEEE Access, vol. 7, 2019, pp. 154922-154934.

[24] P. Yang, D. Wang, Z. Wei, X. Du, and T. Li, "An Outlier Detection Approach Based on Improved Self-Organizing Feature Map Clustering Algorithm," IEEE Access, vol. 7, 2019, pp. 115914-115925.

[25] L. Lyu, J. Jin, S. Rajasegarar, X. He, and M. Palaniswami, "FogEmpowered Anomaly Detection in IoT using Hyperellipsoidal Clustering," IEEE Internet of Things Journal, vol. 4, no. 5, 2017, pp. 1174-1184.

[26] S. Y. Wu, and E. Yen, "Data Mining-Based Intrusion Detectors," Expert Systems with Applications, vol. 36, no. 3, 2009, pp. 5605-5612.