Machine Learning Based Anomaly Detection in Ambient Assisted Living Environments

Ana Cholakoska, Valentin Rakovic,
Hristijan Gjoreski, Marija Kalendar
Faculty of Electrical Engineering and InformationTechnologies
Ss. Cyril and Methodius University
Skopje, R. North Macedonia
{acholak,valentin,hristijan,marijaka}@feit.ukim.edu.mk

Bjarne Pfitzner, Bert Arnrich
University of Potsdam
Hasso Plattner Institute
Potsdam, Germany
bjarne.pfitzner@hpi.de
bert.arnrich@hpi.de

Abstract - Improving the security of the Internet of things is one of the most important and critical issues facing the modern world. With the rapid development and widespread use of the Internet of things, the ability of these devices to communicate securely without compromising their performance is a major challenge. The majority of these devices are limited in power and ability to perform complex computer calculations. This is where anomaly and intrusion detection systems come into play. In this paper, various machine learning algorithms are applied to effectively detect anomalies in such networks. The results obtained show great accuracy and precision (97%), as well as short execution time.

Keywords - Internet of things, machine learning, security, anomaly detection, ambient assisted living

I. INTRODUCTION

The need to facilitate and automate processes is leading to the rapid development of the Internet of Things – part of the fourth industrial revolution. Millions of different devices from multiple manufacturers connected in various ways work together to provide a variety of functions for home, medicine, industry, infrastructure, transportation, and more. However, this diversity poses many problems, mainly related to privacy and security[1,2,6].

Cyber-attacks in such networks are no exception. Depending on the attack surface available, an attack could vary from gaining unauthorized access to a device to power outages, resulting in potentially significant financial and economic losses. The prevailing security threats are especially important in smart homes, utilized for Ambient Assisted Living(AAL).[3]

Using AAL, patients can be monitored around the clock and their clinical outcome may be improved, while reducing costs and optimizing healthcare productivity. Additionally, doctors can easily communicate with patients and have the chance to detect diseases earlier. Nevertheless, the increasing deployment of Internet-connected devices for AAL environments, puts users at significant risk, as personal and health related information become remotely accessible.

Most of the existing studies and frameworks aimed at detecting network anomalies do not explicitly address traffic generated by devices that fall under the Internet of Things[9,11], even fewer work with a relevant data set of an Ambient Assisted Living environment. More recently, machine and deep learning algorithms have been implemented in this area, which have been shown to give good results in detecting such anomalies in networks[10-14].

This paper examines the detection of anomalies in AAL environments by utilizing machine learning algorithms. To the authors' knowledge this is the first work that focuses on ML-based anomaly detection in IoT-centered AAL environments. Section 2 describes the current situation and some of the existing solutions. Section 3 shows the work methodology - selecting the data set and features and processing the data. Sections 4 and 5 provide an overview of the results and a discussion, as well as a conclusion.

II. MACHINE LEARNING FOR ANOMALY DETECTION

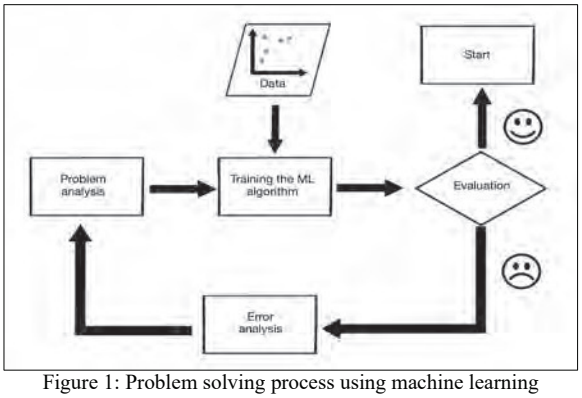

Figure 1: Problem solving process using machine learning

Due to the complexity of modern cyber attacks, the expectations on the modern intrusion detection systems are huge. However, in IoT networks, many specific challenges arise. These devices are constrained, in large numbers and generate heterogeneous data. Therefore, intrusion detection systems that require significant amount of computing or memory usage are not suitable for these networks. So, researchers are inclined to machine and deep learning algorithms to help improve the intrusion detection and prevention process.

Many approaches to tackle these anomaly detection problems have been made and show promising results. Hodo et al.[4], Nobakht[8] and Hossein et al.[18] use multi-layer perception (MLP), logistic regression, support vector machine (SVM) and artificial immune system (AIS) to detect only one type of the following threats: denial of service (DoS), unauthorized access and botnet attacks.

Some approaches, like Golmah[12] and Tanpure[14] propose a hybrid IDS with a classification model to detect abnormal behavior. The first one uses an SVM and the second one uses K-means and Naïve Bayes to group and classify data. It can be seen that such combination of algorithms improves the accuracy of anomaly detection.

However, these approaches don't use data generated from real IoT environments and mostly rely on the KDD dataset[9], which is a dataset that focuses primarily on four types of attacks: DoS, probing, user to root (U2R) and remote to local (R2L). It does not include newer and specific attacks to IoT networks. Also, a machine learning approach for intrusion detection in Ambient Assisted Living environments has not been proposed yet.

III. Methodology

This paper proposes a system for detecting attacks by differentiating anomalies from normal data flow in AAL network traffic. The goal of the system is to detect an anomaly when an attack occurs, i.e., the assumption is that the network behavior will deviate from the normal pattern of behavior and this way the anomaly can be detected[7,15]. To do so, four commonly used machine learning algorithms for real-time anomaly detection: Naïve Bayes (NB), Random Forest (RF), AdaBoost (AB) and K Nearest Neighbors (KNN) have been trained and compared. The classification task is binary, where the anomalies are labeled as 0, while the normal data flow is labeled as 1.

A. Data set selection

To evaluate the approach, a publicly available data set - "IoT Intrusion Dataset 2020" was used, created by a group of researchers from the University of Ontario, Canada[16]. The data set was obtained by simulating network traffic in an AAL network using a smartphone, a security camera, a voice recognition speaker, and several computers. The data set consists of network packets monitored at different time intervals. Table 1 shows the ratio of normal to packets belonging to a particular type of attack. Most anomaly packets refer to the most common attacks that can occur in an IoT network: Mirai botnet, Denial of Service, Scan Port OS, and Man In The Middle (MITM) [17]. Each packet has certain characteristics - package length, source address, destination address, source port, destination port, etc.

B. Data processing and feature analysis

Firstly, preprocessing of the data set was performed. It was purged of zero values, infinite values and incompatible data types. Additionally, feature selection was performed using Random Forest algorithm. The empirical analysis showed the five most informative features and the analysis was continued using only these five features: Flow Duration, Fwd Packet Length Std, Flow IAT Max, Flow Bytes/s, Fwd Packet Length Min.

Table 1: Packet classification

Packet type	Number of packets
Normal	229140
Mirai-UDP Flooding	183554
Mirai-Hostbruteforceg	121181
DoS-Synflooding	59391
Mirai-HTTP Flooding	55818
Mirai-Ackflooding	55124
Scan Port OS	53073
MITM ARP Spoofing	35377
Scan Hostport	22192
Bot	1966

IV. Results and discussion

A. Work environment

The following configuration was used to train and test the refined data set: MacBook Pro with eight-core M1 chip, 8-core graphics card, 16-core Neural Engine and 8 GB RAM. The M1 chip is the first chip system designed by Apple based on the ARM architecture. This chip has the ability to learn 15 times faster than its predecessors, which use Intel architecture. [5]

B. Train and test data set

As can be seen from Table 1, 229140 packets are marked as normal, while the other packets are marked as packages that contain an anomaly (attack). 80% of the data are used in the training of algorithms, while 20% of the data are used for testing and evaluation. In order to be able to compare the performance of the models and find the algorithm that is best for this problem, it is necessary to use several objective statistical indicators - accuracy, precision, sensitivity and F1 score.

Table 2: Results of performed experiments

Algorithm	Accuracy	Precision	Sensitivity	F1 score	Execution time
Naïve Bayes	0.78	0.88	0.62	0.63	**0.57**
Random Forest	0.91	0.93	0.85	0.88	2.25
Ada Boost	0.94	0.93	0.91	0.92	12.49
K Nearest Neighbors	**0.97**	**0.96**	**0.95**	**0.96**	17.0054

C. *Results obtained and discussion*

The experiment was performed ten times for each of the selected algorithms. Table 2 shows the average values of the results obtained.

As can be seen, the RF algorithm, as well as the AB algorithm provide high accuracy and precision (91-94%), but the sensitivity of RF (85%) as well as the F1 result (88%) are lower than those of AB (91%, 92%). In terms of execution time, it can be noted that the AB algorithm runs significantly longer (12.4 seconds) than RF (2.24 seconds) and NB (0.5 seconds).

The KNN algorithm has the longest execution time (17 seconds), but with this algorithm the highest percentages of accuracy (97%), accuracy (96%), F1 score (96%), as well as sensitivity (95%) can be observed. The NB algorithm runs the fastest, but it has the lowest results for all other statistics.

In general, what can be said is that KNN as an algorithm, despite the longer execution time, still gives results that are closest to 100% in terms of accuracy, precision, F1 result and sensitivity. It should also be noted the RF algorithm, which despite the lower sensitivity scores and F1 score, still shows significant accuracy and precision with much shorter execution times. This is especially important in real-time systems.

V. Conclusion and future work

In this paper, research has been done to detect anomalies in IoT-based AAL environment with the help of several machine learning algorithms. Using the characteristics of the data set, the KNN algorithm obtained an accuracy of 97% at an average execution time of 17 seconds. 93% accuracy was also obtained with an average execution time of 2.2 seconds on the RF algorithm.

As future work, these models can be expanded with additional features, in order to improve the previously obtained results. Additionally, other machine learning and deep learning algorithms could be investigated. As a further means of increasing the amount of data and improving the diversity of considered attacks, federated learning could be incorporated.

These approaches are expected to improve the results for classifying anomalies in the Internet of Things in real time.

Acknowledgment

This work has been supported by the WideHealth project - European Union's Horizon 2020 research and innovation programmender grant agreement No. 95227.

References

[1] K. Kimani, V. Oduol, K. Langat, "Cyber Security Challenges for IoT-based Smart Grid Networks", International Journal of Critical Infrastructure Protection, vol.25, 2019.

[2] D. Satria, H. Ahmadian, "Designing Home Security Monitoring System Based Internet of Things(IoTs) Model", Jurnal Serambi Engineering, vol.3, 2018.

[3] V. Venkatesh, V. Vaithyanathan, P. K. Murali, P.R. Chelliah,"A secure Ambient Assisted Living (AAL) environment: An implementation view.",International Conference on Computer Communication and Informatics (ICCCI), 2012.

[4] E. Hodo, X. Bellekens, A. Hamilton, P. Dubouilh, E. Iorkyase, C. Tachtatzis,R. Atkinson,"Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System", 3th International Symposium on Networks, Computers and Communications (ISNCC),2016.

[5] T.Wilder, A.Bender,"Apple unleashes M1", https://www.apple.com/newsroom/2020/11/apple-unleashes-m1/, 2020.

[6] A. Mishra, A. Dixit, "Resolving Threats in IoT: ID Spoofing to DDoS," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-7, 2018.

[7] M. M. Shurman, R. Khrais, A.R. Yateem. "IoT Denial-of-Service Attack Detection and Prevention Using Hybrid IDS", 2019 International Arab Conference on Information Technology (ACIT), pp. 252-254, 2019.

[8] M. Nobakht,"IoT-NetSec: Policy-Based IoT Network Security Using OpenFlow",IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2019.

[9] M. Tavallaee, E. Bagheri, W. Lu, A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1-6, 2009.

[10] M.A. Ferrag, L. Maglaras, S. Moschoyiannis, H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study". Journal of Information Security and Applications, vol.50, 2019.

[11] N. Moustafa, J. Slay, "UNSW-NBI5: a comprehensive data set for network intrusion detection systems (UNSW-NBI5 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, pp. 1-6, 2015.

[12] V. Golmah,"An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM", International Journal of Database Theory and Application, vol. 7, No. 2, pp. 59-70, 2014.

[13] S. A. Hajare, "Detection of Network Attacks Using Big Data Analysis", International Journal on Recent and Innovation Trends in Computing and Communication, vol. 4, issue 5, pp. 86-88, 2016.

[14] S.S. Tanpure, G. D. Patel, Z. Raja, J. Jagtap, A. Pathan. “Intrusion Detection System in Data Mining using Hybrid Approach.”, International Journal of Computer Applications, 2016.

[15] J. Veeramreddy, V.V.R, Prasad, K. M. Prasad, “A Review of Anomaly based Intrusion Detection Systems”, International Journal of Computer Applications, pp.28-35, 2011.

[16] I. Ullah, Q.H. Mahmoud, “A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks”, Goutte C., Zhu X. Advances in Artificial Intelligence, Canadian AI 2020, Lecture Notes in Computer Science, vol 12109, Springer, Cham, 2020.

[17] I. Ahmad, R. Ziar, M. Niazy, S.Khan, “Survey on IoT: Security Threats and Applications”, Journal of Robotics and Control (JRC), vol.2, 2020.

[18] H.R. Zeidanloo, F. Hosseinpour, P. Najafi, “Botnet detection based on common network behaviors by utilizing Artificial Immune System(AIS)”, 2nd International Conference on Software Technology and Engineering, 2010.