The IoT Threat Landscape vs. Machine Learning, a.k.a. Who Attacks IoT, Why do They Do It, and How to Prevent It?

Marek Pawlicki
ITTI Sp. z o.o.
Poznań, Poland *mpawlicki@itti.com.pl*

Bydgoszcz University of Science and Technology
Bydgoszcz, Poland

Aleksandra Pawlicka
ITTI Sp. z o.o.
Poznań, Poland *apawlicka@itti.com.pl*

University of Warsaw
Warszawa, Poland

Mikołaj Komisarek
ITTI Sp. z o.o.
Poznań, Poland *mikolaj.komisarek@itti.com.pl*

Rafał Kozik
ITTI Sp. z o.o.
Poznań, Poland

Bydgoszcz University of Science and Technology
Bydgoszcz, Poland *rkozik@itti.com.pl*

Michał Choraś
ITTI Sp. z o.o.
Poznań, Poland

Bydgoszcz University of Science and Technology
Bydgoszcz, Poland *mchoras@itti.com.pl*

Abstract

Internet-of-Things has been a widely used term, referring to the interconnected ecosystem, built of loosely connected devices, capable of accumulating, processing and transferring data through the heterogeneous network Recently, the IoT's technical, economic and social importance has drastically increased. However, the IoT does not bring advantages only. According to recent studies, vast majority of IoT devices are prone to being attacked, hacked or intruded. If not secure enough, IoT may pose risk to the security of ordinary citizens, and whole industries alike. The paper aims at drawing the current threat landscape in relation to IoT, by examining the threat actors, their motivation and capabilities. Firstly, the specific security goals, context, elements and main challenges to IoT security are discussed. Then, the work collects the actors

that pose the threat to IoT, as well as their motives for attacking IoT. The following part of the paper discusses the various attack taxonomies, and the state-of-the art of the IoT cybersecurity countermeasures and recommendations. Against this background, a novel intrusion detection tool is introduced, and its technical description is provided. When tested on data from a benchmark dataset, the method has already shown promise in performing its tasks.

Keywords: Cybersecurity, IoT, Machine Learning

1. Introduction

Internet-of-Things (IoT) has been a widely used term, referring to the interconnected ecosystem, built of loosely connected devices, capable of accumulating, processing and transferring data through the heterogeneous network [26]. Recently, the IoT's technical, economic and social importance has drastically increased [11]. According to recent studies, vast majority of IoT devices are prone to being attacked, hacked or intruded [14]. What is more, seemingly innocent smart home devices of the ecosystem have been reported to "spy" on users [11]. If not secure enough, IoT may pose risk to the security of ordinary citizens, and whole industries alike [6]. In this paper, the proposition of a component which has the ability to leverage machine learning based classification algorithms to provide real-time network cyberattack detection for data coming from the IoT ecosystem is presented. As will be presented in the paper, the TRENDY (inTRusion in internet of thiNgs DetectIon) component can leverage the entire ML pipeline.

The tool's details are given having first drawn the threat landscape in relation to IoT. This paper is structured as follows: firstly, the specific security goals, context, elements and main challenges to IoT security are discussed. The following part of the paper discusses the various attack taxonomies, and the state-of-the art of the IoT cybersecurity countermeasures and recommendations. Against this background, the TRENDY cybersecurity tool is presented, along with the technical details and its results. Lastly, the paper closes with conclusions.

2. The Emerging Threat Landscape of the IoT

2.1. Security Goals

The discussions on IoT security should be started with the definition of what actually makes an ecosystem secure [28][10]. It has been agreed upon, that the security level of a system may be evaluated by scrutinizing the way it addresses security goals. If some of them are missed, then it becomes possible for a threat actor to tamper with this system. The more of the goals are missing, the more insecure the system is. The goals for cybersecurity comprise of confidentiality, which govenrs if data are disclosed only to appropriate entities and processes; integrity, this goal assures that the data may not be changed without notice, no matter if maliciously or accidentally; availability – in other words, taking care of the system's functionalities and data being always available when expected; accountability – if a system removes the possibility of plausible deniability of actions, then it can be called accountable; authenticity - in an authentic system, all the entities are credible and trustworthy, their identity or other characteristics being verifiable; access control, which ensures that the protected resources may only be accessed by authorized entities [28][10]. The confidentiality, integrity and availability goals are commonly referred to as the "CIA triad" [22].

2.2. Attack Taxonomies

Although the researchers agree about the threat actors who prove dangerous to the IoT ecosystem, they have not arrived at a clear consensus over an attack taxonomy. In the subject literature, several noteworthy taxonomies of the attacks threatening the IoT ecosystems have been

proposed so far. It must be stressed, that a number of taxonomies include physical or cyber-physical attacks as well. However, although they have been mentioned, they are not the focus of this work.

The authors of [29] have proposed dividing the attacks against IoT into the ones related to the Device, Infrastructure, Communication and Service. They argue that the categories are based on the pillars of consumer, commercial and industrial IoT, and real-world IoT attack incidents.

In their work [24], who claim they are the first ones to provide object-based attack categorization in IoT, have divided the attacks on IoT devices into Physical Attacks, Network Attacks, Software Attacks and Data Attacks.

In turn, [28] have proposed dividing attacks according to the affected layer of the IoT system, i.e., Cyber-Physical (CP), Middleware (MW) and Application (APP) layers. The attacks would then be classified according to the identifier based on the relevant security goal: availability (AVAIL), authenticity (AUTH), accountability (ACC), integrity (INT), confidentiality (CONF) and access control (ACL). Using the acronyms, one creates a taxonomy identifier of an attack. For example, an attack on the cyber-physical layer, aiming at the authenticity, would be tagged as CP.AUTH, and so on [28].

Then, [18] divide the attacks on IoT into ones based on: device property (including: Low-end class and High-end class), information damage level (interception/ fabrication/ interruption/ eavesdropping/ modification), location (insider/outsider), strategy (physical/ logical), host-based (user/ hardware/ software), access level (passive/ active), communication stack protocol (physical/ link/ network/ transport/ application) and protocol-based (disruption/ deviation).

The authors of [23] categorize the attacks into three groups: devices and peripherals (brute force, buffer overflow, as well as the rolling code, BlueBorne and Sybil attacks), gateways and internal network (injection attack, MITM, DNS poisoning, replay attacks, wormhole) and cloud servers and control devices (SQL-injection, DDoS, weak authentication), malicious applications and back doors and exploits [23].
Lastly, ENISA has developed the following taxonomy of threats: Nefarious activity/ abuse, Outages, Physical attacks, Disasters, Eavesdropping/ interception/ hijacking, Damage/ loss (IT assets), Failures/ malfunctions [7].

2.3. The State of the Art of Threat Countermeasures and Security Recommendations

The following section demonstrates the state-of-the art of the threat countermeasures, as well as security recommendations, presented in the subject literature.

In their work, [4] outline a set of security guidelines and best practices for securing any Internet-connected device, but especially focusing on the measures “either peculiar to the IoT or especially relevant” to it [4]. They categorize the guidelines into three groups: securing devices, securing networks, and securing the overall system.

A different approach to organizing the countermeasures, described both in [27] and [17], is to divide them according to the affected layer of the ecosystem. The said layers are: the physical layer, the network layer, the processing layer and the application layer. The difference between the authors’ approach is that [17] suggest dividing the IoT threat countermeasures in the same four categories as [27], but offers fewer solutions for each category. Instead, they propose a number of specific, concrete tools or solutions for enhancing security. Herewith, they deem protecting network security to be the most significant challenge, owing to all the vulnerabilities, intricacies and a vast number of various protocols and standards of communication used.

In their work, [23] give a few general recommendations. Namely, they point it out, that the countermeasures have to guarantee integrity, confidentiality and availability. The authors recommend using intrusion detection and prevention systems as the primary countermeasure, combined with the systems preventing data theft or loss [23].

The authors of [16], when discussing the ways of securing IoT, also state that they believe

Intrusion Detection Systems are of fundamental importance [16].

On the other hand, although [2] believes that weak/default passwords are the principal vulnerability, and claims that it would be beneficial to both use strong authentication mechanism and make it impossible for the users to create weak credentials, they also remark that IDS constitute an effective countermeasure against cyberattacks [2].

In their work, [1] discuss four countermeasures helping to enhance security: limiting any unused services or features, which may possibly be used for injecting malicious code, implementing detection mechanism that let only registered users to enter the roots, embedding firewalls, and enabling the encryption of the traffic [1].

In their work, [12] "deduct" the requirements necessary to ensure security, one of their reflections concerning the necessity of constant monitoring and controlling of the ecosystem, to spot any abnormalities as soon as possible [12].

Lastly, along with the threat taxonomy ENISA [7] has presented a list of mitigation measures and practices, aimed at countering the threats; the countermeasures fall into three categories: Policies, Organisational, People and Process Measures, Technical Measures.

The full list of guidelines has been presented in the Annex to [7]; however, the authors yet again stress the significance of monitoring the system in a regular way, searching for malware and integrity errors, conducting audits and security tests, as well as putting proper detection mechanisms in place, and following them with relevant logging, and log correlation and analysis solutions.

In accordance with the state-of-the-art literature, which emphasizes the crucial role of employing intrusion detection systems, an innovative tool was proposed, aimed at detecting network cybersecurity incidents.

3. The Proposed Tool for Intrusion Detection in IoT

In order the tackle the challenge of securing the IoT ecosystem, the TRENDY tool is proposed.

The tool was developed so as to apply to the principles of being modular and scalable. Owing to this, the user is able to build compound machine learning environments. Furthermore, the architecture enables the user to extend the existing components or add new ones.

This was accomplished by implementing the core of the solution in the Scala language, in sync with the Apache Spark library. Such an approach proves effective in deriving all the benefits the Spark platform brings. It also leads to the solution showing high fault tolerance and allows for scalability. The former is managed by means of RDD (Resilient Distributed DataSets) - immutable primary abstraction; the latter is possible owing to the fact that a variety of spark workers can carry out tasks on multiple servers. In addition to this, the architecture's environment is augmented with elements stemming from solutions such as Apache Kafka, Elasticsearch, or Kibana [25].

Apache Kafka is a streaming platform of a distributed nature. It receives, validates, transforms and sends communications between the applications. Apache Kafka plays the role of a broker, which dispatches data coming from various sources to the preprocessors. In turn, preprocessors fulfill the task of preparation of the received input into an output that is fitting for ML. Then, this prepared data may be sent again to the Kafka topic, and applied for further assignments.

Elasticsearch is a distributed, open-source search and analytics engine. Thus, it is capable of collecting substantial amounts of data [9]. The results that Elasticsearch provides are then demonstrated to the user by means of Kibana, a visualisation dashboard.

The abovementioned solution has been presented in Figure 1.

Through a suite of connectors the tool can read and download data from given sources, in the form of a stream. Some parameters, such as the time intervals, or the number of samples to be loaded, may be configured.

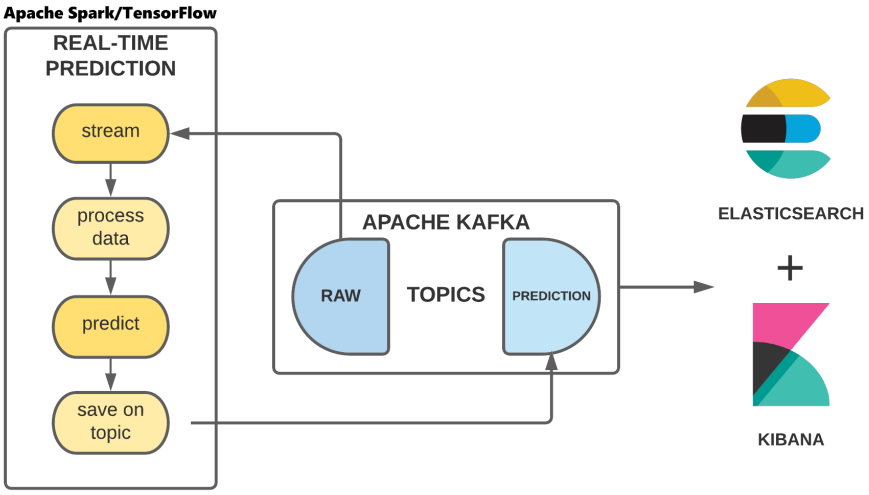

Fig. 1. The pipeline of the TRENDY system

This setup allows to leverage the whole ML pipeline, facilitating adequate handling and preprocessing of data, including data balancing for training data, which is a crucial aspect in IDS [20][13], and using a myriad of different machine learning models, which can be fine-tuned to a specific task. One specific example employable in Network Intrusion Detection is in the use of Deep Neural Networks in the role of supervised classifiers to detect network attacks in NetFlow data coming from a probe or a set of probes set up on routers in the monitored network [21]. The modular nature of the solution allows to utilise TensorFlow, which gives access to the latest advancements in the machine learning domain.

In the ELEGANT project, the IoT ecosystem is where all the inputs come from. The H2020 ELEGANT project aims to develop a novel software solution that addresses key challenges facing IoT and Big Data - interoperability, reliability, safety and security. The whole intrusion detection process was designed for this project with this fact in mind.

Firstly, the specialised network probes collect the network traffic. Then, by means of feature engineering, the data is transformed into relevant feature vectors, which can then be applied for the ML-based intrusion detection. In case the system detects anything of a suspicious nature, an alarm is triggered. Any consequent output is also followed by a comprehensive report being drawn, pointing out annotated results. Lastly, an intrusion detection system of this kind is capable of seizing a wide array of network attacks, due to its operating on both the network and application layers.

4. Results

The effectiveness of network intrusion detection using the TRENDY system is shown in Table 1. The study was prepared with the use of the LITNET 2020 dataset [5], which contains network traffic from the IoT environment. The dataset has 85 unique features, and also contains 12 different network attacks in addition to the normal flow. To study the effectiveness of the solution, a number of metrics for each algorithm have been collected in the table. The experimental stage included the use of algorithms such as Random Forest [3], AdaBoost [8], Naïve Bayes, and Artificial Neural Network (ANN) [15]. The dataset was divided into stratified test and training sets, with the ratio being 70% for training data and 30% for test data. The next step consisted in preparing the data. Incomplete data was removed and fields that were of categorical type were properly transformed. Next, the training set was balanced using the SMOTE method.

Table 1. The performance of each algorithm in detecting anomalies in network traffic originating from the IoT environment; the parameters being Accuracy (ACC), Precision, Recall, F1-Score, Balanced accuracy (BACC), the Matthews Correlation Coefficient (MCC), and Receiver Characteristic Operator (ROC).

Model	ACC	Precision	Recall	F1	BACC	MCC	ROC
Random Forest	**0.97**	**0.97**	**0.97**	**0.97**	**0.8928**	**0.8596**	**0.8928**
AdaBoost	0.97	0.97	0.97	0.97	0.8678	0.8315	0.8678
Naïve Bayes	0.90	0.91	0.90	0.86	0.5417	0.2735	0.5417
ANN	0.96	0.96	0.96	0.96	0.8333	0.7712	0.8333

Finally, the data prepared in this way was fed to the training system. The trained models were finally subjected to verification on test data. For this purpose a number of metrics were prepared to show the effectiveness of the solution. For this particular dataset and setup, Random Forest performed best amongst the examined algorithms.

All in all, now that the work on the technical part of the solution has been completed, there still remain a number of aspects to address. For example, it is necessary to ensure that the solution remains complaint with the General Data Protection Regulation (GDPR) [19].

5. Conclusions

This paper shows that a machine learning-powered tool has the potential to protect IT against cyberattacks. It has sketched the current and emerging IoT threat landscape from the threat actors' perspective, taking into account their motivations and the level of damage they are capable of doing. Then, the various attack taxonomies have been explained, together with the state-of-the-art countermeasures and guidelines, which almost unanimously point at Intrusion Detection Systems as a key element to securing the IoT ecosystem. On the basis of this analysis, the architecture of the TRENDY intrusion detection tool was presented. The method has already shown promise in performing its tasks, and after further work will contribute to safer, more secure Internet of Things.

Acknowledgement

This work is funded under the ELEGANT project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957286.

References

1. Abbas, S.G., Zahid, S., Hussain, F., Shah, G.A., Husnain, M.: A Threat Modelling Approach to Analyze and Mitigate Botnet Attacks in Smart Home Use Case. In: 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE). pp. 122–129. IEEE (dec 2020). https://doi.org/10.1109/BigDataSE50710.2020.00024, https://ieeexplore.ieee.org/document/9343375/
2. Agazzi, A.E.: Smart Home, security concerns of IoT (jul 2020), http://arxiv.org/abs/2007.02628
3. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
4. Corser, G., Fink, G.A., Aledhari, M., Bielby, J., Nighot, R., Mandal, S., Hrivnak, C., Cristache, L.: INTERNET OF THINGS (IOT) SECURITY BEST PRACTICES. Tech. rep. (2017)

5. Damasevicius, R., Venckauskas, A., Grigaliunas, S., Toldinas, J., Morkevicius, N., Aleliunas, T., Smuikys, P.: Litnet-2020: An annotated real-world network flow dataset for network intrusion detection. Electronics **9**(5) (2020). https://doi.org/10.3390/electronics9050800, https://www.mdpi.com/2079-9292/9/5/800
6. Dutta, V., Choraś, M., Pawlicki, M., Kozik, R.: Detection of Cyberattacks Traces in IoT Data. Journal of Universal Computer Science **26** (2020)
7. ENISA: Baseline Security Recommendations for IoT (2017), https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot
8. Freund, Y., Schapire, R., Abe, N.: A short introduction to boosting. Journal-Japanese Society For Artificial Intelligence **14**(771-780), 1612 (1999)
9. Gupta, S., Rani, R.: A comparative study of elasticsearch and CouchDB document oriented databases. In: 2016 International Conference on Inventive Computation Technologies (ICICT). pp. 1–4. IEEE, Coimbatore, India (aug 2016). https://doi.org/10.1109/INVENTIVE.2016.7823252, http://ieeexplore.ieee.org/document/7823252/
10. International Telecommunications Union (ITU).: ITU-TX.1205: series X: data networks, open system communications and security: telecommunication security: overview of cybersecurity 2008. (2008)
11. Iqbal, W., Abbas, H., Daneshmand, M., Rauf, B., Bangash, Y.A.: An In-Depth Analysis of IoT Security Requirements, Challenges, and Their Countermeasures via Software-Defined Security. IEEE Internet of Things Journal **7**(10), 10250–10276 (oct 2020). https://doi.org/10.1109/JIOT.2020.2997651, https://ieeexplore.ieee.org/document/9099839/
12. Kim, H.J., Chang, H.S., Suh, J.J., Shon, T.s.: A Study on Device Security in IoT Convergence. In: 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA). pp. 1–4. IEEE, Jeju Island, Republic of Korea (may 2016). https://doi.org/10.1109/ICIMSA.2016.7503989, http://ieeexplore.ieee.org/document/7503989/
13. Kozik, R., Pawlicki, M., Choraś, M.: Cost-sensitive distributed machine learning for netflow-based botnet activity detection. Security and Communication Networks **2018** (2018)
14. Kozik, R., Pawlicki, M., Choraś, M.: A new method of hybrid time window embedding with transformer-based traffic data classification in IoT-networked environment. Pattern Analysis and Applications (may 2021). https://doi.org/10.1007/s10044-021-00980-2, https://link.springer.com/10.1007/s10044-021-00980-2
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
16. Meneghello, F., Calore, M., Zucchetto, D., Polese, M., Zanella, A.: IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities in Real IoT Devices. IEEE Internet of Things Journal **6**(5), 8182–8201 (oct 2019). https://doi.org/10.1109/JIOT.2019.2935189, https://ieeexplore.ieee.org/document/8796409/
17. Najmi, K.Y., AlZain, M.A., Masud, M., Jhanjhi, N., Al-Amri, J., Baz, M.: A survey on security threats and countermeasures in IoT to achieve users confidentiality and reliability. Materials Today: Proceedings (apr 2021). https://doi.org/10.1016/j.matpr.2021.03.417
18. Nawir, M., Amir, A., Yaakob, N., Lynn, O.B.: Internet of Things (IoT): Taxonomy of security attacks. In: 2016 3rd International Conference on Electronic Design (ICED). pp. 321–326. IEEE, Phuket,

Thailand (aug 2016). https://doi.org/10.1109/ICED.2016.7804660, http://ieeexplore.ieee.org/document/7804660/

19. Pawlicka, A., Jaroszewska-Choras, D., Choras, M., Pawlicki, M.: Guidelines for Stego/Malware Detection Tools: Achieving GDPR Compliance. IEEE Technology and Society Magazine **39**(4), 60–70 (dec 2020). https://doi.org/10.1109/MTS.2020.3031848, https://ieeexplore.ieee.org/document/9290450/
20. Pawlicki, M., Choraś, M., Kozik, R., Hołubowicz, W.: On the impact of network data balancing in cybersecurity applications. In: International Conference on Computational Science. pp. 196–210. Springer (2020)
21. Pawlicki, M., Kozik, R., Choraś, M.: A survey on neural networks for (cyber-) security and (cyber-) security of neural networks. Neurocomputing **500**, 1075–1087 (2022). https://doi.org/https://doi.org/10.1016/j.neucom.2022.06.002, https://www.sciencedirect.com/science/article/pii/S0925231222007184
22. Qadir, S., Quadri, S.M.K.: Information Availability: An Insight into the Most Important Attribute of Information Security. Journal of Information Security **07**(03), 185–194 (2016). https://doi.org/10.4236/jis.2016.73014
23. Rajendran, G., Ragul Nivash, R.S., Parthy, P.P., Balamurugan, S.: Modern security threats in the Internet of Things (IoT): Attacks and Countermeasures. In: 2019 International Carnahan Conference on Security Technology (ICCST). pp. 1–6. IEEE, Chennai, India (oct 2019). https://doi.org/10.1109/CCST.2019.8888399, https://ieeexplore.ieee.org/document/8888399/
24. Sengupta, J., Ruj, S., Das Bit, S.: A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT. Journal of Network and Computer Applications **149**, 102481 (jan 2020). https://doi.org/10.1016/j.jnca.2019.102481
25. Tun, M.T., Nyaung, D.E., Phyu, M.P.: Performance Evaluation of Intrusion Detection Streaming Transactions Using Apache Kafka and Spark Streaming. In: 2019 International Conference on Advanced Information Technologies (ICAIT). pp. 25–30. IEEE, Jinan, China (nov 2019). https://doi.org/10.1109/AITC.2019.8920960, https://ieeexplore.ieee.org/document/8920960/
26. Verma, R., Chandra, S.: A Systematic Survey on Fog steered IoT: Architecture, Prevalent Threats and Trust Models. International Journal of Wireless Information Networks **28**(1), 116–133 (mar 2021). https://doi.org/10.1007/s10776-020-00499-z, http://link.springer.com/10.1007/s10776-020-00499-z
27. Wahab, A., Ahmad, O., Muhammad, M., Ali, M.: A Comprehensive Analysis on the Security Threats and their Countermeasures of IoT. International Journal of Advanced Computer Science and Applications **8**(7) (2017). https://doi.org/10.14569/IJACSA.2017.080768
28. Wustrich, L., Pahl, M.O., Liebald, S.: Towards an Extensible IoT Security Taxonomy. In: 2020 IEEE Symposium on Computers and Communications (ISCC). pp. 1–6. IEEE, Rennes, France (jul 2020). https://doi.org/10.1109/ISCC50000.2020.9219584, https://ieeexplore.ieee.org/document/9219584/
29. Xenofontos, C., Zografopoulos, I., Konstantinou, C., Jolfaei, A., Khan, M.K., Choo, K.K.R.: Consumer, Commercial and Industrial IoT (In)Security: Attack Taxonomy and Case Studies (may 2021), http://arxiv.org/abs/2105.06612