Data Conformity Evaluation: A Novel Approach for IoT Security

Enrico Giulio Maria Verzegnassi
Dept. of Computer, Control, and Management Engineering
"Sapienza" University of Rome
Rome, 00184, Italy
verzegnassi.1772905@studenti.uniroma1.it

Konstantinos Tountas
Electrical and Computer Engineering & Computer Science
Florida Atlantic University
Boca Raton, FL, 33431, USA
ktountas2017@fau.edu

Dimitris A. Pados
Electrical and Computer Engineering & Computer Science
Florida Atlantic University
Boca Raton, FL, 33431, USA
dpados@fau.edu

Francesca Cuomo
Dept. of Information, Elec. and Telecom. Engineering (DIET)
"Sapienza" University of Rome
Rome, 00184, Italy
francesca.cuomo@uniroma1.it

***Abstract*—We consider the problem of attack detection for IoT networks based only on passively collected network parameters. For the first time in the literature, we develop a blind attack detection method based on data conformity evaluation. Network parameters collected passively, are converted to their conformity values through iterative projections on refined L_1-norm tensor subspaces. We demonstrate our algorithmic development in a case study for a simulated star topology network. Type of attack, affected devices, as well as, attack time frame can be easily identified.**

***Index Terms*—Internet of Things, security, L_1-norm, data conformity, tensors, tensor decomposition, principal-component-analysis.**

I. Introduction

A growing number of diverse physical devices are being connected to the Internet at an unprecedented rate realizing the idea of Internet of Things (IoT). IoT enabled physical objects may sense the environment, and perform tasks together by communicating with each other to coordinate decisions [1]. An IoT network is not a simple wireless sensor network; it is a dense integration of the virtual and the real world where communication between devices and humans takes place. In a sense, it can be considered as an interwoven medium of heterogeneous networks of different sizes making up a large global network. Applications such as transportation, healthcare, industrial automation, and smart homes are just some domains which IoT networks will enable or enhance [2]. IoT shifts from functionality to connectivity and data-driven decision making. New requirements and constraints are introduced on the network, on the physical communication channel, on the end device hardware and software [3]. End devices are tailored to their assigned tasks to have the lowest possible size, cost, and energy consumption [4]. As a consequence, frequently the hardware has low resources in terms of computation power, memory, battery, transmission range, throughput, and delay. At the same time, for continuous secure IoT operations key requirements are fault-tolerance, self-healing against compromised nodes and attacks, energy efficiency, scalability, and flexibility for heterogeneous network formation over a wide area [4].

This new heterogeneous networking environment creates significant challenges in applying known security schemes to the IoT [3], [4]. Most of the state-of-the-art security schemes are tailored to common networking standards and architectures, in contrast with the goals defined in [5]. On the contrary, a single IoT network could comprise of smaller sub-networks utilizing different networking architectures and standards, as well as end devices with different computation, communications, and security capabilities and requirements.

Recently the research community started investigating the use of intrusion detection systems (IDS) for the IoT networks, in order to tackle some of the security challenges and successfully identify attacks and affected nodes in wide wireless networks. Specifically, a complete analysis of an IDS for WSNs is presented in [6]. The authors utilize the OSI model and geographical attack propagation to classify attacks. In [7], the authors analyze different IDS technologies and provide useful guidelines for potential IDS applications in WSNs. Descriptions of different WSN attacks are presented in [4], where the authors provide detailed discussion about the general concept of WSN security and its challenges.

Machine learning techniques such as principal component analysis (PCA) for intrusion detection are introduced in [8]. The authors present a dimensionality reduction and classification system designed to detect malicious activity and intrusions. Moreover, the work in [9] utilizes machine learning ideas to capture global knowledge of the traffic patterns in the wireless networks in order to identify intrusion attacks.

In this paper, we present for the first time in the literature

The work of D. A. Pados was supported by the Schmidt Family Foundation, Boca Raton, FL.

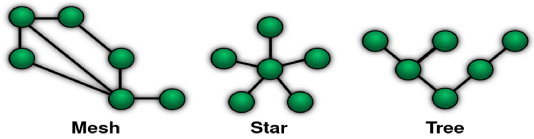

Fig. 1: Typical wireless network architectures.

a blind tool for attack detection based on network parameter conformity evaluation. The system utilizes only end devices' network parameters passively collected by the base station (BS) or by dedicated devices and calculates how conforming the parameters across the end devices are in time. Conformity is calculated iteratively by successive projections of the dataset on carefully calculated subspaces that capture the relationship between devices, network parameters, and time. Attacks on end devices result in changes in the network parameters, translating to low data conformity of these devices at the specific time intervals. The proposed solution does not require edge device collaboration for attack detection and works independently of the IoT network architecture and implementation.

The rest of this paper is organized as follows. Section II discuses security concerns in different IoT network architectures and network parameters that may want to include in our dataset. Section III describes the proposed data conformity evaluation algorithm. Performance evaluation of the algorithm on simulated network data is presented in Section IV. A few concluding remarks are drawn in Section V.

II. Problem Statement

IoT security solutions should preserve user privacy, confidentiality, data integrity, availability, and non-repudiation. However, active edge device collaboration for security purposes is challenging as it leads to increased energy consumption, software and hardware resources, and maintenance requirements. Moreover, the solutions should be agnostic to the network architecture and edge device type. Different network topologies are presented at Fig. 1. In a typical star topology, the BS serves as the sink where all edge devices report to. Hence, the BS is able to gather data regarding the devices' state and features, such as received signal strength (RSS) and packet delivery ratio (PDR), organically, without the need of explicit edge device collaboration. This may enable anomaly detection based only on parameters gathered locally. In other end device topologies, such as mesh or tree networks, it is challenging for an edge device to obtain such information.

In a network where the edge devices communicate directly to each other, collecting RSS and PDR values for each individual link is difficult without external help. To solve this problem, researchers assume the presence of specialized devices (agents) collecting data from the network [10]. Typically, these agents are devices with high hardware capabilities able to gather and report network parameters, which due to the device geographical separation and the nature of the network's architecture would not have been possible to obtain otherwise. Two types of agents are considered, passive and active. Passive agents are not part of the IoT network and are able to only acquire data about their nearby environment and end devices without interacting with the rest of the network. Active agents are able to communicate with the rest of the network, request and exchange information and are considered as "super" powerful edge nodes.

Different attacks affect different network parameters in a distinct way depending on the attacker's goal and the IoT network architecture and topology. In our studies, we consider parameters that an agent or a BS is able to obtain and calculate through normal network operation. These parameters include PDR (the number of packets delivered over a specific time interval), RSS (the average received signal strength over a time interval), the number of different recipients' addresses, the distance between the agent/BS and the edge device, and the edge devices' latitude and longitude. Edge device battery consumption is another important parameter affected by many attacks, however, a passive agent is not usually able to retrieve it, and therefore an active agent should be considered.

Next, we present two important attacks on an IoT network that can be identified through the aforementioned parameters.

- Jamming attack: The main goal of this attack is to disrupt the communication of the end devices close to the attacker. The jamming attack can be divided into four types [11]: constant, random, deceptive and reactive. The signal disruption involves all the devices close to the jammer and the effect for the victims is a substantial decrease in the number of packets delivered while the RSS at the BS or the agent remains relatively high. Identification of a jamming attack can be carried out from the PDR, RSS, percentage of attacked victims, and the distance between victims values.
- Sybil: In such an attack, a single node forges multiple identities delivering packets with wrong information and deceiving the intended receiver. Demirbas et al. [12] show how it is possible to detect the Sybil attack by using multiple "agents" collecting RSS data of the affected nodes. In fact, if two or more nodes are uniquely identified, and/or located at the same point, they are considered forged, thus a Sybil attack can be detected [12], [13]. The device's location can be estimated through RSS, message delay (time-of-flight), or angle-of-arrival (AOA) localization techniques [14].

III. Data Conformity Evaluation Algorithm

For our dataset, we define the tensor $\mathcal{X} \in \mathbb{R}^{D \times L \times N}$ where the columns correspond to end devices, rows to the parameters, and the third dimension to time. The goal of the algorithm is to convert the original data tensor $\mathcal{X}$ to a tensor $\widetilde{\mathcal{W}} \in \mathbb{R}^{D \times L \times N}$ which contains the conformity value of each tensor element in the $[0, 1]$ range (Fig. 2). A security analyst, human or machine, can then analyze the $\widetilde{\mathcal{W}}$ tensor and infer which devices are affected by different attacks at different time instances.

Next, we describe how we calculate the conformity of the data through iterative L_1-norm tensor decompositions. Without loss of generality, we present the developed methodology with

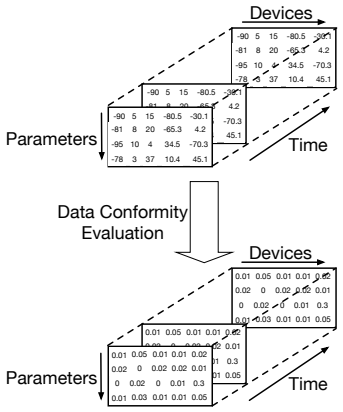

Fig. 2: The original dataset is converted to element-by-element conformity values. The lower the value the less conforming the element.

respect to the first mode of our tensor dataset $\mathbf{X}_{(1)} \in \mathbb{R}^{D \times LN}$. First, we calculate R_1 principal components of the first mode $\mathbf{X}_{(1)}$ denoted as $\mathbf{Q}_1^{(1)} \in \mathbb{R}^{D \times R_1}$ by solving the following optimization problem

$$\mathbf{Q}_1^{(1)} = \underset{\substack{\mathbf{Q} \in \mathbb{R}^{D \times R}, \\ \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_R}}{\operatorname{argmax}} \left\| \mathbf{X}_{(1)}^T \mathbf{Q} \right\|_1. \tag{1}$$

The optimization problem in Eq. (1) calculates robust subspaces by explicit L_1-norm projection maximization. The resulting principal components are called L_1-norm principal components, which have been proven to be extremely resistant to misbehaving data [15]. We utilize this property to calculate the conformity of our data by defining the L_2-norm distance between each column of $\mathbf{X}_{(1)}$ and the calculated subspace $\mathbf{Q}_1^{(1)}$ as

$$d_{1,i}^{(1)} = \left\| \left(\mathbf{I}_D - \mathbf{Q}_1 \mathbf{Q}_1^T \right) \left[\mathbf{X}_{(1)} \right]_{:,i} \right\|_2^{-1}, \quad i = 1, 2, \ldots, LN. \tag{2}$$

We expect small $d_{1,i}^{(1)}$ values if $d_{1,i}^{(1)}$ is an "outlier" and large if it is a nominal data column. Therefore, the conformity of each data column of $\mathbf{X}_{(1)}$ can be measured as the L_2-norm distance from the calculated subspace, i.e.,

$$\mathbf{w}_1^{(1)} = \begin{bmatrix} d_{1,1}^{(1)} & d_{1,2}^{(1)} & \ldots & d_{1,LN}^{(1)} \end{bmatrix}_{LN \times 1}^T. \tag{3}$$

The vector $\mathbf{w}_1$ is then converted to a tensor, where each column i is weighted by the same coefficient $d_{1,i}, i = 1, 2, \ldots, LN$.

$$\mathcal{W}_1^{(1)} = \text{tensorization} \left(\mathbf{w}_1^{(1)} \circ \mathbf{1}_{D \times LN} \right) \tag{4}$$

where the "tensorization" operation converts the mode-n matrix back to the original tensor form. After calculating the tensors $\mathcal{W}_1^{(1)}$, we repeat the same process for the two other modes to obtain $\mathcal{W}_2^{(1)}$ and $\mathcal{W}_3^{(1)}$. Then, we combine in an additive fashion to form the final data conformity tensor

$$\mathcal{W}^{(1)} = \alpha_1 \mathcal{W}_1^{(1)} + \alpha_2 \mathcal{W}_2^{(1)} + \alpha_3 \mathcal{W}_3^{(1)} \tag{5}$$

where $\alpha_k \in \mathbb{R}^+, k = 1, 2, 3, \sum_k \alpha_k = 1$ are weights corresponding to the prescribed importance of each dimension of the dataset (for example $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$, if all treated equal). The weight tensor $\mathcal{W}^{(1)}$ is then normalized

$$\widetilde{\mathcal{W}}^{(1)} = \frac{\mathcal{W}^{(1)}}{\left\| \mathcal{W}^{(1)} \right\|_1}. \tag{6}$$

The final weight tensor $\widetilde{\mathcal{W}}^{(1)}$ contains the data conformity value of each individual element enabling element-wise conformity evaluation of the data. Then the L_1-norm subspace of the first mode may be refined by

$$\mathbf{Q}_1^{(1)} = \underset{\substack{\mathbf{Q} \in \mathbb{R}^{D \times R_1}, \\ \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{R_1}}}{\operatorname{argmax}} \left\| \mathbf{Q}^T \left(\mathbf{X}_{(1)} \circ \widetilde{\mathbf{W}}_{(1)}^{(1)} \right) \right\|_1, \tag{7}$$

where $\circ$ is the element-wise (Hadamard) product. The same holds true for the other modes. The data conformity values can be iteratively refined until numerical convergence of the data conformity tensor $\widetilde{\mathcal{W}}^{(l)}$ is observed. In Table I, we present the complete pseudo-code of the algorithm for a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$.

Algorithm 1 Data conformity evaluation through L_1-norm tensor decomposition

Input: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, ranks $R_1, R_2, R_3 \in \mathbb{Z}^+$, and weights $\alpha_1, \alpha_2, \alpha_3 > 0$

Output: $\widetilde{\mathcal{W}}$

for k = 1, 2, 3 **do**
  $\mathbf{Q}_k^{(0)} = \underset{\substack{\mathbf{Q} \in \mathbb{R}^{I_k \times R_k}, \\ \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{R_k}}}{\operatorname{argmax}} \left\| \mathbf{Q}^T \mathbf{X}_{(k)} \right\|_1$
  $M_k = \prod_{i=1, i \neq k}^{3} I_i$
end for, $l = 0$
while convergence criterion is not met **do**
  for $k = 1, 2, 3$ **do**
    $d_{k,i_k}^{(l)} = \left\| \left(\mathbf{I}_{I_k} - \mathbf{Q}_k^{(l-1)} \mathbf{Q}_k^{(l-1)^T} \right) \left[\mathbf{X}_{(k)} \right]_{:,i_k} \right\|_2^{-1}, \forall i_k = 1, 2, \ldots, M_k$
    $\mathbf{w}_k^{(l)} = \begin{bmatrix} d_{k,1}^{(l)} & d_{k,2}^{(l)} & \ldots & d_{k,M_k}^{(l)} \end{bmatrix}_{M_k \times 1}^T$
    $\mathcal{W}_k^{(l)} \leftarrow \text{tensorization}_k \left(\mathbf{w}_k^{(l)} \circ \mathbf{1}_{I_k \times M_k} \right)$
  end for
  $\widetilde{\mathcal{W}}^{(l)} = \frac{\sum_{k=1}^{3} \alpha_k \mathcal{W}_k^{(l)}}{\left\| \sum_{k=1}^{3} \alpha_k \mathcal{W}_k^{(l)} \right\|_1}$
  for $k = 1, 2, 3$ **do**
    $\mathbf{Q}_k^{(l)} = \underset{\substack{\mathbf{Q} \in \mathbb{R}^{I_k \times R_k}, \\ \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{R_k}}}{\operatorname{argmax}} \left\| \mathbf{Q}^T \left(\mathbf{X}_{(k)} \circ \widetilde{\mathbf{W}}_{(k)}^{(l)} \right) \right\|_1$
  end for, $l = l + 1$
end while

TABLE I: Data conformity evaluation through L_1-norm tensor decomposition

IV. Case Study

For testing our algorithm, we choose two attacks which are most relevant for IoT networks, namely constant jamming and

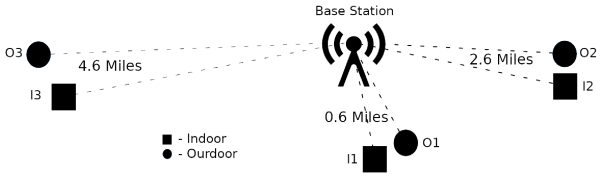

Fig. 3: Simulated network: Three end devices located outside, and three nodes located inside buildings.

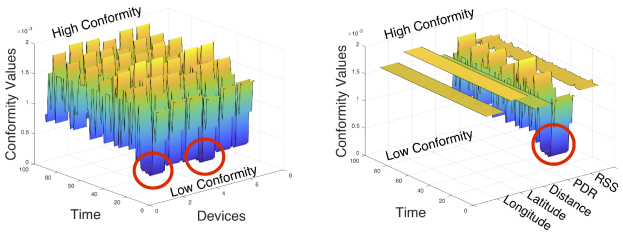

Fig. 4: (a) PDR conformity values across devices. (b) Parameter conformity across time for attacked device 4. The red circles indicate the non-conforming parts of the dataset.

sybil. The simulated network comprises of six end devices and a BS in a star topology shown in Fig. 3. Three of the devices are located inside buildings and the rest are located outside in known locations. The devices utilize the LoRaWAN protocol and LoRa physical layer. They are programmed to sent at least two packets to the base station every minute. The sampling of the devices' state is done passively by the base station every minute. For calculation of the RSS parameter we utilize the well known Friis transmission equation [16] assuming that each device and the BS are equipped with a single omnidirectional antenna and are transmitting at the 900 MHz frequency band.

A. Detection of Constant Jamming Attack

Fig. 4-(a) depicts the calculated PDR conformity across all end devices. We observe that between the $20-40$th slabs end devices 1 and 4 are affected by an attack, as their PDR conformity indications are much larger compared to the rest. Fig. 4-(b) depicts the conformity of all parameters of end device 4. It is clear that the device is affected by an attack, as all parameters have a high conformity, while PDR has low conformity.

B. Detection of Sybil Attack

During the Sybil attack an attacker forges messages of a random victim for a period of 20 minutes (or 20 slabs). This results in the BS to misinterpret the received packets, and assign them to the attackers location. The parameters latitude, longitude, distance, RSS are affected by this attack. In Fig. 5-(a) depicts the longitude conformity value across all devices. It is clear that device 2 is under attack as the longitude conformity is higher compared to the rest. Moreover, on the same time interval, the distance parameters is non-conforming

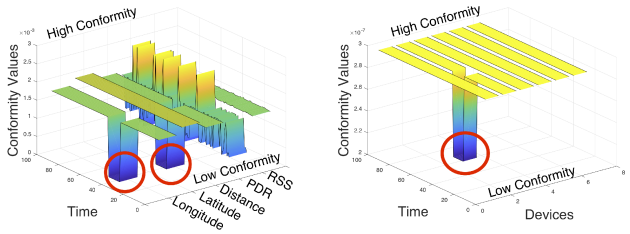

Fig. 5: (a) Distance conformity across edge devices. (b) Parameter conformity across time for attacked device 2. The red circles indicate the non-conforming parts of the dataset.

as well, which means that the node is under the Sybil attack (Fig. 5-(b)).

V. Conclusions

We considered the problem of attack detection for IoT networks based only on passively collected network parameters. For the first time in the literature, we developed a blind attack detection method based on data conformity evaluation. Network parameters collected passively, are converted to their conformity values through iterative projections on refined L_1-norm tensor subspaces. We demonstrated our algorithmic development in a case study for a simulated star topology network. Type of attack, affected devices, as well as, attack time frame can be easily identified.

References

[1] F.A. Alaba, M. Othman, I.A.T. Hashem, F. Alotaibi, "Internet of Things security: A survey", *Elsevier J. Netw. Comput. Appl.* 88 (2017) 10–28.
[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, 2010.
[3] D. Mocrii, Y. Chen, P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security", *Journal of Internet of Things*, vol. 1-2, pp. 81-98, Sept. 2018.
[4] K, Tamil, D. Sridharan, "Security Vulnerabilities In Wireless Sensor Networks: A Survey", *Journal of Information Assurance and Security*, vol. 5, pp. 31-44, 2010.
[5] M. Chernyshev, Z. Baig, O. Bello and S. Zeadally, "Internet of Things (IoT): Research, Simulators, and Testbeds," in *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1637-1647, June 2018.
[6] V. Garcia-Font, C. Garrigues, and H. Rifá-Pous, "Attack classification schema for smart city WSNs", *Sensors*, vol. 17, Apr. 2017.
[7] O. Can, O. K. Sahingoz, "A survey of intrusion detection systems in wireless sensor networks", *IEEE Communications Surveys Tutorials*, vol. 16, issue 1, pp. 266-282, May 2013.
[8] H. H. Pajouh, R. Javidan, R. Khaymi, A. Dehghantanha, and K. R Choo, "A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks", *IEEE Transactions on Emerging Topics in Computing*, Nov. 2016.
[9] S. Suthaharan, "Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning", *Performance Evaluation Review*, vol. 41, issue 4, pp. 70-73, 2014.
[10] T. Bhattasali, R. Chaki, and S. Sanyal, "Sleep Deprivation Attack Detection in Wireless Sensor Network", arXiv preprint arXiv:1203.0231.
[11] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The Feasibility of Launching and Detecting Jamming Attacks in Wireless Networks", *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc)*, pp. 46-57, Urbana-Champaign, IL, USA, May 25 - 27, 2005.

[12] M. Demirbas and Y. Song, "An RSSI-based Scheme for Sybil Attack Detection in Wireless Sensor Networks," *Proc. of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pp. 564-570, 2006.

[13] S. T. Patel, and N. H. Mistry, "A Review: Sybil attack detection techniques in WSN", *In Proc. 4th Int. Conf. on Electronics and Communication Systems (ICECS)*, Coimbatore, India, 24-25 Feb. 2017.

[14] P Sarigiannidis, E Karapistoli, and A. A. Economides, "Detecting Sybil attacks in wireless sensor networks using UWB ranging-based information", *Expert Systems with Applications*, vol. 42, issue 21, pp. 7560-7572, Nov. 2015.

[15] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, "Efficient L1-norm principal-component analysis via bit flipping", *IEEE Trans. on Signal Processing*, vol. 65, pp. 4252-4264, Aug. 2017.

[16] H.T. Friis, "A Note on a Simple Transmission Formula', *Proceedings of the IRE*, vol. 34 , issue 5, pp. 254-256, May 1946.