Traffic Characterization to Provide Trust for Internet of Things Devices

Evandro L. C. Macedo  
High-Speed Networks Laboratory  
Federal University of Rio de Janeiro (UFRJ)  
Rio de Janeiro, Brazil  
evandro@ravel.ufrj.br

Luís F. M. de Moraes  
High-Speed Networks Laboratory  
Federal University of Rio de Janeiro (UFRJ)  
Rio de Janeiro, Brazil  
moraes@ravel.ufrj.br

Abstract—The Internet of Things (IoT) is paving the way for the development of Cyber-Physical Systems (CPS), the next step of the Internet evolution, which will allow the development of several new systems and applications. Likewise, urged by the adoption of 5G and Beyond networks, the massive, ubiquitous spread of interconnected IoT devices has increased the vulnerability of data and related applications in an unprecedented manner. If the security of any component in such systems gets compromised, affecting its trust with respect to others, an associated data leak may cause serious threats to privacy, material losses, and even put people’s lives at risk. In this paper, we present IoT devices’ traffic characterization to provide trust values to enable secure communications among such devices. We develop experiments using a real IoT dataset to demonstrate the feasibility and effectiveness of our proposal. Considering that complementary features between blockchain technology and information theory trigger a great potential for research and innovation, the key idea of the contribution consists in modeling trust using a two-level approach, which is based on a distributed-ledger (at the high level), and a relative entropy measure (at the low level). The results show the feasibility of our approach.

Index Terms—blockchain, entropy, IoT, security, trust

I. INTRODUCTION

As emergent paradigms for networks evolution, the fifth network generation (5G) [1], and the Internet of Things (IoT) [2] expand the network communication horizons to provide any type of smart object (things) with ultra-reliable low-latency communication (URLLC). A plethora of benefits can be obtained with the growing adoption of IoT, for instance, in the areas of smart cities, healthcare, intelligent transportation systems, Industrial IoT, and many others based on IoT devices [3]. Particularly, IoT enables a noteworthy data acquisition within the ecosystems previously mentioned, which can afford improvements for various decision-making processes.

In this context, it is of paramount importance to provide IoT systems with security. Indeed, a security breach can lead to a data leak that may cause serious privacy threats, bring about material losses, and even jeopardize people’s lives. Therefore, new requirements and challenges need to be considered in the design and development of IoT systems and applications [4], especially in terms of security.

In this paper, we consider our trust model presented in [5]–[7] to provide additional analysis using an IoT traffic characterization to clarify how our approach works and verify the feasibility and effectiveness of the proposed approach with experiments using a real IoT dataset [8]. Such dataset is composed of several raw packets which include flow information for each pair of devices (source and destination) with traffic from both licit and malicious devices included. The data consists of IoT devices traces including cameras, smart lights, activity sensors, and health monitors, with some of them presenting abnormal behaviors.

The key contribution of the proposed trust model consists in combining both characteristics of Low Level (network perspective) using the relative entropy of the incoming traffic of an IoT device, and High Level (application perspective) with a distributed-ledger-based approach (blockchain), to compose a comprehensive trust metric, capable of capturing changes in the behavior of devices and isolating those who present unexpected misbehavior.

The rest of this paper is organized as follows. In Section II we describe the problem addressed in this paper and discourse related works. Section III presents the trust components and our trust model. In Section IV the experimental analysis and result discussion are presented. Finally, Section V concludes the paper and foresees future work.

II. PROBLEM STATEMENT AND RELATED WORK

Security aspects is admittedly a major challenge in IoT [9]–[16], due to the heterogeneity between the multitude of components and platforms IoT interconnects, the resource-constrained devices, and the wireless communication technologies, which are inherently more vulnerable. In particular, the problem of assigning trust metrics to IoT devices is of prime importance and is still considered a challenge [17], [18]. Thus, this work aims to explore IoT trust-building solutions and propose an approach that ensures the trustworthiness of devices throughout their communication [19].

To build trust among entities, many works in the literature propose approaches based on Blockchain (BC). Authors in [20] present an in-depth survey about the integration of BC
and IoT, discuss the insights of this new paradigm, and explore an integrated architecture called Blockchain of Things and its application possibilities among many domains, e.g., supply chain, health care, internet of vehicles, among others. Fortino et al. [21], [22] designed a framework in which every IoT device is associated with a software agent capable to exploit its social attitudes to cooperate as well as to form complex agent social structures. The authors consider the reputation aspect by using a BC implementation and devices can use network services according to their reputation provided by BC.

Authors in [23] focus on fog computing level to offer a bi-directional trust management system for secure offloading and fog-to-fog collaboration, allowing a service requestor to determine the trustworthiness of a service provider and vice-versa before initiating a connection. They use fuzzy logic to aggregate trust obtained using quality of service, quality of security, social relationships, and past reputation metrics.

Hongjun et al. [24] use Information Theory to build trust among devices. They represent the relationships with a directional graph and compute the entropy of the capability of a device in performing an action. This way, they can detect malicious devices in the network. We also consider Information Theory in our work, but with a different perspective focusing on the network level instead of the application level.

In [25], authors present a survey on trustworthiness and dependability in IoT systems and propose a framework to provide trustworthiness at the data level for mist and fog-based IoT systems. Their framework provides data trustworthiness to ensure a continuous and uninterrupted operation of IoT data flow. They also discuss challenges and trade-offs related to data trustworthiness in IoT and present data flow proposals for four different possible stages according to their framework, namely, thing, mist, fog, and cloud stage.

The aforementioned works emphasize the importance and relevance of building trust-based approaches to provide security in the communication among IoT devices. Nevertheless, the reviewed works did not present any multiple perspective solution in terms of considering not only the application level, but also the network level. Such approach is important in order to provide a comprehensive trust solution that can gather more aspects into consideration when building a trust metric than with just one-perspective approach. Therefore, in this paper, besides presenting a trust model that combines Blockchain and Information Theory techniques, the key contribution of our work is the double perspective of both application level and network level. Hence, we can provide a more comprehensive trust metric that can deal with the particularities of IoT devices traffic patterns and the specificities of applications.

III. TRUST MODEL

To model trust for IoT, we need to know which information is necessary to compose a trust metric. According to the model proposed in [6], a receiver IoT device needs to build an initial trust to enable communication, since it does not know the sender IoT device previously. Given the initial trust, the receiver should dynamically adjust its trust in the sender accordingly to the network behavior of the sender (more trust if the behavior is as estimated, less trust otherwise). Then, the trust value should not last endlessly, but instead, be decreased over time, given that, after the last contact, the receiver does not know if the sender was compromised or if it presents any unexpected (potentially malicious) behavior.

An illustration of how the proposal works can be seen in Figure 1. Initially, in the illustrated scenario, two entities (Alice and Bob) do not know each other, having little or no information needed to infer an initial trust value to start communicating. The operating steps are as follows:

- **Step 1:** Each IoT device queries the other’s identity in a Blockchain infrastructure that stores the identities of all devices, which is in the High Level of our approach;
- **Step 2:** Once a minimum initial trust is established, the communication can normally start over the Internet, either through a Thing Tier to Cloud Tier communication, or a Thing Tier to Edge Tier communication;
- **Step 3:** As the communication between Alice and Bob happens, the Low Level takes place. Alice calculates the relative entropy of Bob’s traffic (and vice versa) and uses this information to adjust the value of trust in Bob over time. If Bob starts behaving abnormally, this will negatively affect trust and may cause communication to terminate if it decreases beyond a previously established threshold. If the communication is over, a temporal component reduces the trust value until it reaches the point where Alice and Bob will have to query the Blockchain again and restart the whole process.

We consider an IoT system composed of three tiers, namely, the Things tier, the Cloud tier and an intermediate Edge tier [26] as Figure 1 depicts. Such an organization is driven by the recent paradigm of Edge Computing [27], [28], which aims to move computing, processing, and storage resources to the edge of the network, rather than centralizing them in remote cloud data centers. With such approach, an infrastructure is created that provides lower latency for applications, as compared to the cloud. Edge devices can range from switches, routers, or

---

1In our case, an abnormal behavior means any traffic pattern that diverges from the estimated traffic distribution.
base stations to smart gateways or micro data centers and they usually have more limited resources than cloud devices, but possess far more capabilities than IoT devices.

Considering stationarity, let \( X_{ji} \) represent the data rate (in Bytes per second – Bps) that a device \( j \) receives from a device \( i \). If the traffic generated by a device is below the capacity of the data link connecting that device to the network, then the traffic is equal to the throughput. \( X_{ji} \) is considered to be a random variable that can assume values in the sample space \( S_{ji} = \{0, \Delta, 2\Delta, 3\Delta, ..., R_{ji}^{\text{max}}\} \), where \( \Delta \) is a positive integer and \( R_{ji}^{\text{max}} \) is the maximum received data rate. Hence, the distribution of \( X_{ji} \) is given by \( P[X_{ji} = x] \approx p_{X_{ji}}(x), \ x \in S_{ji} \).

Assume \( TR_{ji} \) as the trust of device \( j \) in device \( i \), defined in the interval \([0, 1] \). The \( TR_{ji} \) is initially defined based on the reputation of the device stored in the Blockchain. This component gives a trust value based on the number of confirmations a device receives from base stations to smart gateways or micro data centers and they usually have more limited resources than cloud devices, but possess far more capabilities than IoT devices.

In our model we consider a proportional temporal decay.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

In this section, we evaluate the potential of our approach on translating the network traffic behavior of IoT devices into a meaningful trust metric. We perform experiments using both real and synthetic datasets, considering boundary behaviors to test the effectiveness of our approach. It is worth mentioning that we are not simulating the trust metric, but we are actually calculating it in a realistic scenario since we are using real data obtained from a real IoT dataset. All dynamism that is typical of such a context is reflected in traffic traces. For example, the connectivity disruption due to mobility causes zero traffic to be received by a device.

Results presented in this section use the dataset found in [8]. Traffic traces from a smart-campus environment compose this dataset with over 20 IoT devices, including cameras, smart lights, activity sensors, and health monitors. These traces include raw packets (pcap) and flow information over a period of 3 weeks. In our experiments, we considered a period of one day of the dataset and extracted the traffic in bytes/s from flows of each pair of devices according to the tuple (Source IP, Destination IP) by summing the number of bytes transmitted in one second. In this way, we have a ratio of one sample per second.

Figure 2 depicts the original traffic history over time an IoT device \( i \) sent to device \( j \). We also considered a compromised version of the same traffic sample with an anomalous traffic (50 KBytes/s) insertion from the instant 13000 sec to 17300 sec (Figure 3). We use this compromised version of the traffic to analyze the behavior of our approach under such anomalous circumstances, which may characterize malicious traffic behavior.

We run experiments considering the traffic between any two devices identified inside the dataset through their respective flows. The experiment consists in playing the traffic values obtained from the dataset within the interval of measurement for each pair of devices. There are some relevant assumptions regarding the implementation to be highlighted:

- Taking into account the typical resource constraints of IoT devices, the results were obtained considering a sliding window size of 600, which contains the traffic values used in the traffic distribution estimation. This is to assess how the tiny sliding window performs to embrace devices that are only capable of processing at most such a window size;
- For the sample distribution of traffic we consider a fixed amount of bins (10) and a fixed maximum value of traffic (100 KBytes/s);
- The value for the estimated traffic, used to compare with the value for the actually received traffic, is calculated by decreasing the trust value from the moment devices stop communicating until they reach a threshold. When trust value falls below the threshold, devices will need to return to the first case of trust establishment, i.e., devices will need to obtain a minimum trust from Blockchain again.
using a Kalman Filter, since it closely tracks the received traffic and does not require a lot of resources;

- When the communication is established for the first time between two devices, only the component at the High Level (application-based) actuates, obtaining the reputation of the device in the community;
- When there is traffic (i.e., the incoming throughput is greater than zero and less than the transmission capacity), only the Low Level components (network-based) actuate to change the trust value;
- When there is no traffic (i.e., the incoming throughput is zero), only the temporal component (also at the Low Level) actuates by constantly decreasing the trust value according to a predefined rate (e.g. \(-0.1\) trust/s);
- We consider a threshold of 0.9 as a minimum trust value necessary for a device to communicate with another device. This means that a device with trust below the related threshold is not allowed to communicate.

A. Analysis of the proposal

For the following results, we observe the impact of using network characteristics and application-level information to compute the trust metric. Primarily, all available samples of one day were used, namely 21600 samples, which corresponds approximately to a 6-hour collection period. Then, we analyze an interval of the samples to provide an analysis closer to what an IoT device could perform (given its resource constraints).

We analyze the distribution of the samples through the following histograms, in which the x-axis corresponds to the amount of incoming traffic in an IoT device in bytes/s, and the y-axis corresponds to the relative frequency observed according to the specified number of samples. The x-axis varies between the minimum and maximum values found within the sample range considered in each case. It was considered 10 bins, i.e., 10 intervals of the same length that section the x-axis in 10 equal parts. We also provide the goodness of fit of distribution taking into account a Wakeby distribution to approximate the sample distribution of traffic.

In the plots in Figures 4 and 5, all available samples were used to calculate the relative frequencies. Most of the traffic values are relatively low or zero, although we have values with a high flow rate as well. Comparing Figure 4 with Figure 5, we can observe the insertion of anomalous values in the trace under analysis through the relative frequency increase close to the interval of 50 KBytes/s.

Since IoT devices are meant to use a sliding window to compute trust, by analyzing a smaller range of samples within the trace, we can observe the anomalous behavior does not appear until the moment in which it occurs is reached. In Figure 6, the range considered is that of the first 1000 samples and the anomaly does not appear. Conversely, considering the interval between 15500 and 16500 of the samples (Figure 7), the histograms have significantly changed.

To observe the change in the goodness of fit of the Wakeby
distribution on the sample data we compute the Kolmogorov-Smirnov (K-S) statistic, which presents 0.413 for the original trace and 0.362 for the modified trace. For the smaller range in which the anomalous traffic behavior occurs, it is noticeable that the Wakeby distribution does not fit well (Figure 7), with a K-S statistic 0.330, which means that the parameters of the distribution have changed. Bearing this in mind, the device receiving the traffic can calculate the distance between the sample distribution of the observed data and the estimated distribution (of which traffic follows a Wakeby) previously learned. This distance is given by calculating the relative entropy, also known as Kullback-Leibler divergence, which we discussed in Section III.

Figure 8 depicts the results of trust values calculated for the original trace. The trust increases as soon as the traffic pattern begins to stabilize and varies according to traffic behavior. Then, it shows a trust stabilization from around the 7000 sec to the end of the trace. In Figure 9 though, when the inserted anomalous behavior is reached (at 13000 sec), the trust value begins to change its pattern. During this period, as the trust values are below the established threshold (0.9), the device is not allowed to communicate. In this case, it has to rely on its reputation, which is provided by the High Level component. Then, as the modified period passes, trust values start to increase again. These changes are produced by the variations on traffic distribution over time, captured by the relative entropy and temporal components (the Low Level of our proposal), which could not be perceived from an application-level perspective.

Given these results, our approach can capture changes in the traffic patterns of the devices, which helps to compose a trust metric. Therefore, the approach prevents malicious devices from successfully communicating with other devices on the network.
Fig. 9. Trust values obtained with our trust model using the modified version of the traffic from the dataset [8]

V. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we present an analytical model for trust in the context of IoT. We gather characteristics of Low Level (network perspective) and High Level (application perspective) to compound a meaningful trust metric capable of capturing network traffic behavior changes. The results obtained using real and modified traffic traces show that our approach behaves according to the expectations, including in extreme behaviors caused by spikes purposely inserted into the real dataset. Therefore, we show the effectiveness of our approach in capturing network behavior changes, adjusting trust according to that, and protecting the licit devices from malicious ones.

For future work, we plan the model extension considering Artificial Intelligence aspects to improve the learning of new traffic behaviors IoT devices might present. We also envision a real deployment considering devices virtualization with digital twins in a Multi-access Edge Computing context, and a real blockchain implementation to provide efficiency metrics and results from a real deployment.

ACKNOWLEDGMENT

We would like to acknowledge the support from Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Rede-Rio (the state academic backbone network), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), which partially funded this study.

REFERENCES