J. E. Rubio, R. Roman, and J. Lopez, "Integration of a Threat Traceability Solution in the Industrial Internet of Things", *IEEE Transactions on Industrial Informatics*, vol. 16, 2020.
http://doi.org/10.1109/TII.2020.2976747
NICS Lab. Publications: https://www.nics.uma.es/publications

Integration of a Threat Traceability Solution in the Industrial Internet of Things

Juan E. Rubio, Rodrigo Roman, Javier Lopez
Department of Computer Science, University of Malaga,
Campus de Teatinos s/n, 29071, Malaga, Spain
{rubio,roman,jlm}@lcc.uma.es

Abstract

In Industrial Internet of Things (IIoT) scenarios, where a plethora of IoT technologies coexist with consolidated industrial infrastructures, the integration of security mechanisms that provide protection against cybersecurity attacks becomes a critical challenge. Due to the stealthy and persistent nature of some of these attacks, such as Advanced Persistent Threats, it is crucial to go beyond traditional Intrusion Detection Systems for the traceability of these attacks. In this sense, Opinion Dynamics poses a novel approach for the correlation of anomalies, which has been successfully applied to other network security domains. In this paper, we aim to analyze its applicability in the IIoT from a technical point of view, by studying its deployment over different IIoT architectures and defining a common framework for the acquisition of data considering the computational constraints involved. The result is a beneficial insight that demonstrates the feasibility of this approach when applied to upcoming IIoT infrastructures.
Keywords: Intrusion, Detection, Traceability, IIoT, Opinion, Dynamics, Industry.

1 Introduction

Critical infrastructures, such as industrial systems, have traditionally worked in isolation from external networks such as the Internet. However, modern technologies (e.g. Big Data, Cloud Computing) are increasingly being integrated in such rigid environments in order to provide additional benefits, such as service automation and cost reduction. One of these technologies is the Industrial Internet of Things (IIoT). The conventional Internet of Things is envisaged to interconnect day-to-day objects to the Internet and the cloud, so as to monitor their behaviour and manage them easily. When it comes to the industry, the IIoT has the ultimate goal of providing benefits to the production chain such as predictive maintenance and production optimization. Yet the inherent heterogeneity associated to the IIoT (multiple technologies developed by

diverse actors) and its connectivity capabilities also widens the attack surface, increasing the risk of cyber-security attacks.

In such a unstable scenario, security must be consolidated to keep up with the progress of integrating novel technologies such as the IIoT. In this sense, traditional solutions to perform an early detection of intrusions are just a first line of defense and are not enough to anticipate the stealthy actions of the most sophisticated threats. In turn, Opinion Dynamics is a novel approach that is specifically conceived for this purpose, by proposing a multi-agent collaborative system that permits to trace down the actions and impact of a sophisticated attack, returning useful information to accurately address and repair the affected resources. It was originally proposed to be integrated in a traditional industrial control system, but its openness to include multiple indicators of detection makes it able to be integrated in a IIoT-based infrastructure. In this paper, we revisit this approach and conduct a concise study of the applicability of this promising solution to the particular context of the IIoT, while also addressing all its structural and computational constraints.

The remainder of this paper is organized as follows: Section 2 presents the technologies considered in the IIoT, surveys the most used intrusion detection and traceability techniques in this context and introduces the Opinion Dynamics approach. In Section 3 the applicability study is carried out considering all its stages of implementation. Based on these findings, a conceptual example of the integration of the Opinion Dynamics system over a IIoT network with the theoretical simulation of an attack is presented in Section 4. Finally, the conclusions drawn are presented in Section 5.

2 Industrial Internet of Things

2.1 Industrial Internet of Things Technologies

At present, there are multiple actors that are defining the technologies that comprise the IIoT [1]. Such actors include various standardization groups (e.g. IETF) and several consortia (e.g. the Industrial Internet Consortium (IIC) [2] and the Platform Industrie 4.0 consortium [3]). As a result, the IIoT technology ecosystem is very heterogeneous, ranging from standards that originated from specific industry verticals to protocols that were designed for general-purpose use. These technologies provide all the necessary components to build a functional IIoT infrastructure: from hardware and software platforms to communication technologies at the lower and upper layers of the networking stack.

From a **hardware perspective**, a "thing" in the IIoT can be any sensing or actuating device that interacts with the physical world and can be accessed through the Internet protocol suite – either directly or indirectly. These entities range from existing industrial devices enhanced with additional networking capabilities and high-level services (e.g. Programmable Logic Controllers (PLCs) equipped with the MQTT protocol) to sensor/actuator devices equipped with wireless connectivity (e.g. WirelessHART sensors forming a capillary network).

The capabilities of these devices in terms of memory and computational power is also very heterogeneous, ranging from constrained nodes to more capable devices.

From a **software perspective**, there are various reference architectures whose goal is to provide additional services beyond the basic exchange of data, including operation, management, business logic, and security. The most important reference architectures are the Industrial Internet Reference Architecture (IIRA) developed by the Industrial Internet consortium [2], and the Reference Architectural Model Industrie 4.0 (RAMI4.0) developed by the Platform Industrie 4.0 consortium [3]. Although as of 2020 there are no complete instantiations of these reference architectures, the functionality of some of their components is being verified through the use of testbeds. Moreover, certain major industry players, such as Siemens [4], already provide basic IIoT solutions.

As for the communication technologies and protocols, they can be classified into two categories: lower layer protocols and upper layer protocols. **Lower layer protocols** are deployed under the network layer (IP), and in the context of the IIoT all protocols make use of a wireless transmission channel (cf. [1]). These protocols can be classified as *Wireless Personal Area Networks (WPAN)* (e.g. IEEE 802.15.4, Bluetooth), *Wireless Local Area Networks (WLAN)* protocols (e.g. IEEE 802.11), and *Cellular Networks and Low-Power Wide-Area Networks (LPWAN)* protocols (e.g. 4/5G, SigFox). In the context of industrial networks, the main difference between these technologies is the location of the gateway that connects the wireless network with the industrial network: In WPAN and WLAN, gateways can be deployed and controlled at the industrial premises, while in cellular networks data must first traverse the telecommunications network before reaching the specific industrial network that consumes the information – which can be located on premises or in the cloud. Also, most WPAN networks make use of subsets of the IP standards (e.g. 6LowPAN) or proprietary protocols (e.g. WirelessHART).

Upper layer protocols are deployed over the transport layer (TCP or UDP), and allow the exchange of information in a shared data structure between participants. The most important upper layer IIoT protocols, as defined in [1] and [2], are *Messaging and data-oriented protocols* like MQTT (which focus mainly on publish-subscribe mechanisms), lightweight RESTful *Web Services*, and *Specific frameworks* such as OPC-UA (an evolution of the OPC specification that provides better semantic modelling) and OneM2M (a service layer that provides efficient communication between application endpoints).

2.2 Traceability of attacks in Industrial and IIoT scenarios

As with traditional IT systems, IIoT deployments can be attacked by malicious adversaries, which could generate serious operation disruptions in critical infrastructures. In this context, intrusion detection systems (IDS) become a necessary defense layer to detect potential attacks against these infrastructures. Even if the field of IDS for IIoT technologies is not as developed as the field of IDS for traditional industrial ecosystems (cf. [5]), there is still a plethora of detection

approaches. Some of these detection mechanisms focus on the integration of *signature-based IDS* and Deep Package Inspection (DPI) technologies [6], which try to find specific patterns in the network frames. Other *anomaly detection systems* implement various machine learning techniques, aiming to detect instances of data (exchanged from IIoT devices) that do not belong to a learned class (i.e., a model that has been trained and validated).

Besides, there are several IDS specifically designed for IIoT deployments that benefit from the unique characteristics of industrial networks (e.g. deterministic operation procedures) compared to general IT networks [7]. According to the state of the art (cf. [8]), these intrusion detection procedures mainly focus on the analysis of the communication patterns and the protocols states to identify a deviation from a previously created specification. This leads to two main detection strategies: *specification based anomaly detection* and *physical state dynamic estimation*.

In the first strategy, *specification based IDS*, human experts build a model that describes the legitimate system behavior (e.g., protocols, programs, operations) to latter compare it with the current state to detect anomalies. Some examples of this approach include [9], where an advanced metering infrastructure is modelled to represent a legitimate activity profile at various levels, and [10], where the specification is at protocol-level to model the Modbus TCP communication patterns. The second strategy, *physical state dynamic estimation*, complements the first strategy by modeling the physical dynamics of the operations performed in the production chain. For example, in [11], the authors propose a resilience framework for cyber-physical systems which permits to describe physical domains mathematically. Other examples include [12] and [13], which models the physical constraints of a power grid infrastructure and a water distribution network, respectively.

Regardless of the detection strategy used in the industrial premises, IDS only pose a first line of defense, and further post-incident analysis of the generated evidences (e.g., alarms, network events) and raw traffic must be conducted all across the network to anticipate the effects of sophisticated and persistent attacks such as Advanced Persistent Threats (APTs) [14]. This is carried out by traceability and advanced correlation mechanisms, which provide information of the overall network health status and facilitate the deployment of accurate response measures based on the threat evolution. This has been mostly addressed in traditional corporate environments, by means of proactive techniques (evidences are analyzed as incidents occur) and reactive techniques (evidences are studied once the events occur). Among the former, [15] proposes a framework for flow-based analysis of network traffic in near real time to detect APTs in Cloud Computing. Also, in [16], researchers present a security framework for the analysis of high volumes of traffic to identify data exfiltrations and suspicious activities in TCP/IP networks. Some other approaches conduct advanced analytics with the outputs of external IDS. For example, in [17], researchers propose an approach entitled TerminAPTor, a theoretical supervision system capable of linking multiple information flows from classical IDS. In [18], the authors propose MLAPT, a machine learning-based system to detect and predict

APT attacks by correlating the outputs of different detection methods. As the rest of approaches, it is experimentally validated in a corporate infrastructure (using a dataset of attack scenarios against a campus network).

As for industrial ecosystems, the introduction of increasingly dynamic topologies and the growing range of security threats in the IIoT and Industry 4.0 complicate the process of information acquisition [19]. Moreover, to the best of our knowledge, all existing traceability approaches are designed for generic IT networks, and have not explicitly discussed how they could be implemented and validated using real attacks. Therefore, as no traceability solution exists that takes into account the IIoT context, it is the main motivation of this paper to provide a first step in this area.

2.3 An intrusion detection and traceability framework based on Opinion Dynamics

After reviewing some of the most representative methods for intrusion detection in IIoT environments, this section presents the Opinion Dynamics detection and traceability framework, as the analysis of the applicability of this existing framework to IIoT ecosystems forms the core of the paper. The framework was originally described in [20], where the approach was presented in theory. Its authors then enhanced the attack model [21] and the event correlation process [22]. In practice, this framework was implemented in a realistic industrial setting in [23], and the authors demonstrated its applicability of the approach to a Smart Grid ecosystem in [24].

The framework consists of a cooperative multi-agent system whose (virtual) agents correspond to all existing industrial devices deployed in the network. The algorithm followed by the elements of this framework is presented in Figure 1, and its comprised by six stages. In **S1**, *data retrieval setup*, the system extracts the outputs of multiple anomaly detection mechanisms, vulnerability scanners or SIEM systems. In **S2**, *agents creation*, all data associated to a particular entity or device is assigned to its corresponding virtual agent. Note that raw data not extracted from existing IDS, such as network traffic, can be used to obtain additional features (e.g. traffic volume, type of connections established) in **S3**, *Feature extraction*.

In **S4**, *feature selection and opinion formation*, Each agent i combines all available data into an opinion $x_i(t)$, which shows the opinion (i.e. anomaly value) of the agent at a given time t (i.e., the security state of its monitored node, measured from 0 to 1). For this task, different models can be applied to weigh each feature depending on the current security scenario and the anomalies sensed. The evolution of these opinions over time is considered in **S5**, *Correlation of opinions*. In this phase, all opinions evolve by taking into consideration the opinions of the surrounding agents $x_j(t)$ and a weight w_{ij}. In order to facilitate this process, in the current incarnation of this framework this correlation is executed in a central system. All opinions evolve using the following expression:

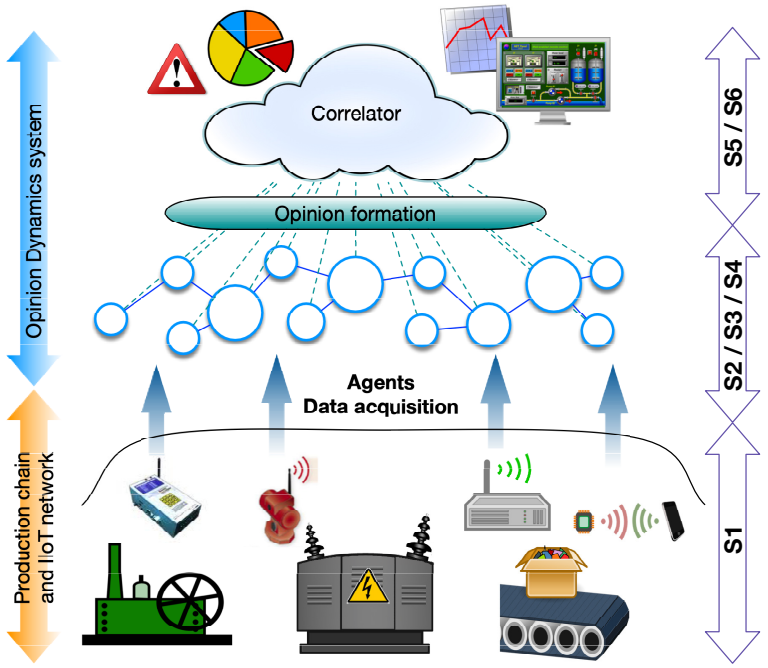


Figure 1: Stages of the Opinion Dynamics framework in a IIoT network

$$x_i(t+1) = w_{i1}x_1(t) + w_{i2}x_2(t) + ... + w_{in}x_n(t)$$

As a result of this correlation, it is possible to extract additional indicators in **S6**, *Computation of indicators*. For example, all opinions can be grouped into clusters at any given time, providing a representation of the segments of the network that are being affected by existing attacks. Moreover, a global health indicator can also be calculated from the aggregation of all opinions. This opinion model can be enhanced by taking into account other parameters such as the criticality of the monitored resource, its historical events, or the persistence of the detected attacks.

As a result of its design, this framework has provided various contributions to the current state of the art on IDS in industrial ecosystems, such as i) circumventing the heterogeneity of IDS solutions by combining various solutions to provide protection at all levels, ii) facilitating the accommodation of new technologies and business scenarios due to the adaptable nature of the framework, and iii) easing the traceability of attacks and the precise application of response procedures thanks to the dynamic partitioning of the elements of the network and the correlation of events.

Nevertheless, further research is necessary to fully realize the Opinion Dynamics framework, as there are several open questions to be solved. For example,

whether using a centralized entity is a feasible solution in all scenarios, or how to precisely instantiate these agents (e.g., IDS, anomaly detection mechanisms) on a physical infrastructure whose criticality may restrict the modifications of hardware and software. Additionally, the potential overhead introduced in the communications, or the provisioning of parallel network interfaces to gather and analyze network traffic are other open issues that we aim to resolve in this paper. Especially, we aim to successfully apply the Opinion Dynamics approach in the IIoT domain, taking into consideration the constraints mentioned in the Introduction. More specifically, we will study the precise instantiation of the algorithm, making more emphasis on the earliest stages – as they revolve around the integration of the algorithm with the IIoT network at low level.

3 Applicability of the Opinion Dynamics System to the IIoT Scenario

3.1 Feasibility of Data Retrieval (S1)

To start devising the integration of the Opinion Dynamics framework over an IIoT scenario, the main question that arises is the nature of the information that can be collected by the detection system. As stated previously, we must provide the agents (regardless of where they are executed) with data of interest about the state of the resource they are monitoring, as to finally output a single – but aggregated – value of anomaly, that represents its opinion (stage S4). This process requires of data that is retrieved in **stage S1**, either from raw information extracted from the low layer and high layer protocols or from outputs of IDS solutions such as the ones described in section 2.2. Here, we will especially focus on the former, as the existence of IIoT IDS already proves their feasibility as inputs to the framework. In general, the information that can be processed by agents include, but are not limited to:

- **Network parameters:** involves two kinds of information, related to the topology and the state of the network, to infer the presence of anomalies via traffic analysis (by comparing the current value with the one learned in normal conditions):

    1. A physical network mapping that contains every pair of devices connected through a communication channel (in form of a graph, with the address of every node within the topology). This can be easily determined from the number of packets per protocol and recipient, which helps to tag frequent and non-frequent communications.
    2. Quality of Service indicators: they inform about the reliability of connections by means of metrics like the delay time from one node to another, the bandwidth experienced and the packet loss ratio in connection-oriented protocols.

- **Communication information:** it implies the analysis of the payload contained within the exchanged packets and their frequency, which includes low-level commands issued from one source to its destination (e.g., control commands to actuators), as well as quantitative values from operations (e.g., readings from sensors). The former allows to detect suspicious actions potentially performed by compromised devices, while the latter permits to create a statistic model to later identify deviations in the values exchanged.

Going back to the early stages of the algorithm, the method for extracting these features from the traffic in a IIoT network is highly dependent on the wireless transmission channel used, its particular deployment architecture, and the application endpoint where data is consumed (which is presumed to be the central correlator). The aim with stage S1 is to seamlessly gather the aforementioned network information without interfering with the operations of the production chain (i.e., additional computation and delays) and, whenever possible, without introducing extra physical equipment. This imposes several challenges, such as inferring a low-level network mapping out of the application data received by upper layer protocols (e.g., when only a gateway is visible for the industrial segment as an interface to the IIoT subnetwork) or estimate indicators through a parallel communication channel when the primary one is inaccessible (e.g., in third-party cellular networks).

Consequently, we must start by studying the amount and quality of data that can be potentially collected from the IIoT network given a specific configuration. For the sake of clarity, we define the concept of OT cell as a subsection of the entire industrial infrastructure where the same underlying wireless technology is implemented. Thus, according to the classification of lower layer protocols described in Section 2.1, we can draw some conclusions about the network parameters that can be obtained:

- **WPAN networks.** Both classic Bluetooth and the low-energy specification (the latter featuring the creation of a large-scale mesh of devices) support connectivity at IP-level in certain nodes within a network, acting as bridges between the industrial domain and the sensors at field level. As for IEEE 802.15.4 devices, gateways (e.g., coordinators in a Zigbee network) often centralize the retrieval of data from the lower layers of the industrial architecture. Therefore, the network-related information that is possible to extract in a OT cell of this kind is the one retrieved by the gateway that interconnects it with the upper levels of the infrastructure. This usually implies that the original information exchanged by sensors/actuators using these lower layer protocols is translated by the gateway into common industrial standards such as ModbusTCP, thereby losing granularity when studying the precise topology and QoS indicators. Consequently, we have three alternatives: 1) to deploy a capillary network that captures and relays the missing information through an auxiliary network interface (introducing hardware in exchange); 2) to manually provide

the network mapping information at low level and establish the relationship with high level packets (lacking the QoS information); 3) to rely on this aggregated data and carry out a deep analysis of high level packets to infer the network mapping.

- **Wireless Area Networks (WLAN).** IEEE 802.11 standards, and in particular the latest 802.11ah standard, facilitate the creation of IIoT networks where a large number of devices need to cover wider areas. In contrast with WPAN networks, this is achieved with a higher power consumption, which enables the use of the IP protocol in all devices to cover areas of up to 1000m in a single hop. In addition, Relay Access Points are used to extend the connectivity to Access Points (APs), that transparently deliver the field level information to the industrial network, without any routing between the endpoint and the gateway. From the data acquisition perspective, this means that the network mapping and QoS indicators are easily obtained by capturing and analyzing the exchanged traffic packets.

- **Cellular networks.** When collecting low-level information in Cellular Networks, the amount of packets that can be captured decreases dramatically due to the presence of a public telecommunication network that processes all the traffic before it is consumed in the industrial network. Thus, it is not possible to obtain QoS data while packets are relayed through the multiple hops of the external infrastructure. Plus, the network mapping must be inferred at logical level, by capturing application level traffic and accounting for every source-destination pair within the industrial premises. This scarce amount of information increases when an edge paradigm is leveraged (e.g., fog computing or mobile edge computing) or when some of the cellular network infrastructure assets are controlled privately by the company, instead of an external provider.

Wireless transmission channel	Network parameters accesible	
	Network mapping	QoS
WPAN (IEEE 802.15.4, Bluetooth)	Through an additional capillary network, analysing high-level data from the gateway or manually	From the IT/OT network to the gateway only
WLAN (IEEE 802.11)	Yes, all data	Yes, all indicators
Cellular Networks and LPWAN	Logical network mapping, unless external telecomm. infrastructure or edge network resources are monitored	end-to-end indicators, unless external telecom. infrastructure or edge network resources are monitored

Table 1: Network parameters collected from the different IIoT cells

Table 1 summarizes the different methods for collecting low-level network information in each IIoT cell. Still, stage S1 does not depend only on the information provided by lower layer protocols – it also also revolves around gathering

information about the communications at application level, as explained before. This can be classified into two classes: information about the production chain from the field devices, and control commands issued from the IT section to the industrial process. As for the former, the process of extracting the measured data from sensors is relatively straightforward, depending on the upper layer protocol used to exchange data:

- In asynchronous message protocols and publish-subscribe mechanisms such as MQTT or AMQP, the entity in charge of running the detection algorithm should be registered as subscriber to receive the measurements from the broker (i.e., the intermediate gateway).
- In RESTful architectures like CoAP or HTTP, the sensors readings would be accessed by means of an API (published by a CoAP server executed on an intermediate gateway or embedded in the own device on the field).
- In frameworks such as OPC-UA and OneM2M, the retrieval of data requires additional analysis of how it is generated and consumed by endpoints, since they respond to abstract specifications of communication interfaces between services and components that are integrated in specific domains. It usually implies reading values from a common server that exposes a friendly API under a unified data model.

It is worth noting these communication channels very frequently use encryption measures to ensure the confidentiality of data (e.g., CoAP is built on top of DTLS). This makes it necessary that the entity that retrieves data from devices and executes the detection algorithm is allowed to access the exchanged data and comply with the system access control policy.

On the other hand, we also should be able to retrieve the precise set of commands that are issued from the managerial level of the industrial network, as explained before. According to the architecture of a IIoT-based control system, this implies filtering the operations executed by a PLC, which is hierarchically placed on top of an IIoT cell and ultimately issues commands to sensors/actuators (potentially using intermediate IIoT gateways). These devices can operate with a large range of protocols, ranging from open source standards like ModbusTCP or Ethernet/IP to private alternatives such as S7 from Siemens. In this case, accessing to the commands executed requires the development of dissectors for the particular protocol, which exceeds the scope of this paper. However, as there are numerous solutions available in the market that especially focus on the analysis of these standards [5], it is possible to use external IDS results as inputs for our system.

3.2 Opinion Dynamics in IIoT Networks (S2-S4)

In this section, we introduce the design of the rest of the stages of the Opinion Dynamics algorithm in IIoT networks. Note that we do not analyze stages S5 and S6, as these stages are independent from the underlying infrastructure once

all necessary information (e.g. opinions) is available. For this particular instantiation, we will make use of the information extracted in section 3.1, without resorting to external systems (i.e. existing IDS systems). Note that, due to the nature of the framework, such IDS can be integrated anytime.

The virtual agents created in **stage S2** deal with the processing of data retrieved in S1 and the features extracted from S3. From a physical point of view, this firstly means that the central correlator that executes the Opinion Dynamics System must establish a communication channel with every IIoT cell that is being monitored, in order to gather the network parameters (e.g., a link to the gateway in a WPAN or to the AP in a WLAN network). Likewise, it must be able to access the interfaces where data is published (e.g., the API in a CoAP based network). Then, from a logical perspective, this information in bulk is divided and assigned to virtual agents created by the correlator.

These agents, according to the theoretical approach [21], are threads in charge of individually monitoring the security of an IIoT device within the topology to subsequently derive an opinion, following a 1:1 relationship between devices and agents. Equivalently, an agent receives the traffic (containing data and commands) that is exchanged by its assigned device, as well as the QoS indicators of every connection that it shares with the rest of neighbours. At this point, the physical network mapping conducted in S1 is essential for the central correlator to make such assignment of information. Nevertheless, as discussed before, the knowledge about the physical topology is not always accurate, due to the presence of intermediate gateways that aggregate data from a mesh of constrained devices and hinder the retrieval of network parameters. In this case, when the actual mapping cannot be determined by any of the methods presented in Table 1, we can assume the existence of agents that encompass a set of multiple devices.

Once agents are created and provided with the information that they need to process, they perform an extraction and selection of features from that data in **stage S3**. These features refer to variations in certain magnitudes or indicators, which evidence anomalies suffered as a consequence of an attack. Some examples of features applicable to IIoT networks are:

- Number of connections established and devices accessed
- Traffic load (total number of packets exchanged)
- Type of communication protocols used
- Delay experienced in every communication channel
- Ratio of lost/corrupted packets
- Frequency and type of commands issued
- Precise data values transmitted by sensors

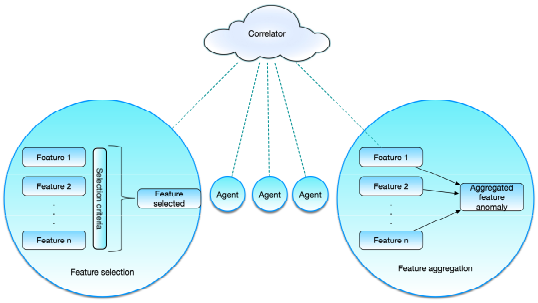
These features are monitored periodically (as often as the Opinion Dynamics is executed to visualize the latest changes in the network). A model is created

to represent the behavior of each one so that it is updated in every period. Even though diverse alternatives could be proposed for formalizing this model, here we conceptually propose a simple but accurate approach, which is internally used in commercial IDS: in the case of quantitative values (e.g., number of packets), the average is calculated. As for discrete features (e.g., devices accessed or protocols used), the model is represented with the set of occurrences for each value (e.g., number of packets sent to a given recipient or using a certain protocol) and their corresponding average. Either way, the values obtained for each feature are compared in each period with the existing model, which is assumed to reflect the behavior of the system in normal conditions (therefore, a initial phase of training is assumed). As a result, the standard deviation provides a value of anomaly for quantitative features. In discrete ones, the value of anomaly can be determined by analyzing the individual deviation in the number of occurrences. This way, the extraction of features would be complete for each agent.

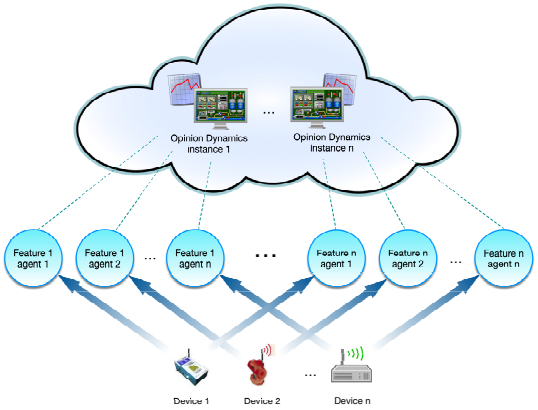
All of these features are closely related to intrinsic network aspects of the devices monitored. A future work could involve the analysis of host-based parameters in the own IIoT devices as a source of anomaly for the opinion computation. For instance, the usage of CPU and memory, the processes running, and others. This would require the integration of capillary networks that retrieve such information from the OT cells or using external detection systems. This is possible due to the adaptable nature of the Opinion Dynamics framework, which is open to include all kinds of features.

The opinion formation in **stage S4** is the last stage before the correlation of anomalies and analysis of detection results. The opinion of each agent is formed at this point by deriving a single value from the set of anomalies sensed in each feature, which implies making a selection or aggregation. Diverse policies could be applied and compared, being the easiest to *select the feature* whose anomaly value is the maximum as the opinion for a particular agent. This would make the overall results of the opinion dynamics system very sensitive to changes, since a singular feature from the complete set of indicators measured by an agent could influence a whole neighbourhood of agents and raise risk alarms indicating the presence of a threat. Still, this approach could be recommendable in highly critical infrastructures where a fine grained auditing is needed. An alternative to selection is the *aggregation of features*, using the average of anomalies sensed for all the indicators considered (as long as they are not zero), for instance. However, the drawback of this approach is that greater anomalies measured in important features would be occulted to the correlator due to the aggregation with lower anomalies in other features. In this case, a weighted average of features would be interesting.

Lastly, there is one more way to implement this stage and avoid the loss of detail as a consequence of a selection or aggregation. It consists in *conducting a Opinion Dynamics correlation per feature considered*, so that multiple instances of the detection algorithm are executed in the centralized entity, where each one concerns on a specific indicator; in that case, the correlator would take the anomalies in each feature as individual opinions for all the Opinion Dynamics instances (equivalently, each device would have an agent per feature monitored).



(a) Feature selection and aggregation



(b) Multiple Opinion Dynamics instances for features

Figure 2: Alternatives for the opinion formation in S4

As a result, it would be possible for a security administrator to visualize the state of connections, delays, protocols, etc. with a deeper level of detail. All these three alternatives are summarized in Figure 2 and shown in the next section through a simple example.

After the formation of opinions in all agents of the network, they can be correlated and analyzed in **stages S5 and S6** using the Opinion Dynamics algorithm to visualize the clusters of agents that expose the same degree of anomaly measured in their surroundings. This information is useful for computing health indicators for diverse areas and carry out a precise analysis of the historical data to draw conclusions about the attack pattern and predict future actions, as explained in [21].

4 Case study

After the study of the applicability of the Opinion Dynamics in the IIoT, this section focuses on showing the benefits of a conceptual deployment of this approach

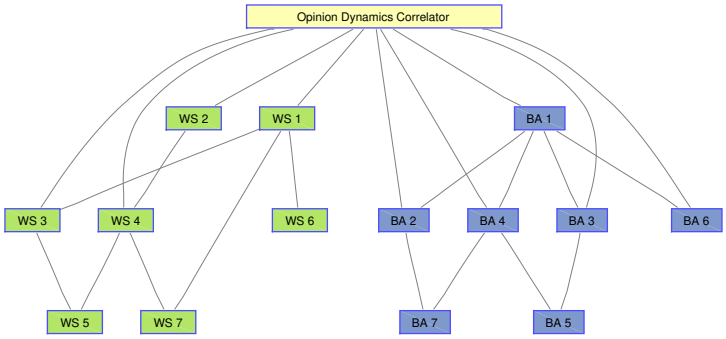


Figure 3: Example of network composed by two IIoT cells, using the Watts–Strogatz (WS) and Barabási–Albert (BA) model

by means of a theoretical study case. Additionally, the three aforementioned alternatives for conducting stage S4 are discussed. In order to achieve these goals, we will follow the same methodology of the original publication [21] to represent the attacker model and its evidenced anomalies using graph theory.

The formalization of the network is explained in the following: firstly, we define a graph that represents the physical interconnection of the Opinion Dynamics system with the multiple IIoT cells that are present in the infrastructure, from which data is retrieved. For this purpose, the authors of the original paper used a power-law network random distribution to simulate a traditional industrial infrastructure, which can be punctually connected to the Internet. In this case, we leverage the Watts-Strogatz [25] and the Barabási-Albert model [26]. Both distributions permit to simulate the topology of an IIoT cell, being the former used for producing graphs with small-world properties [27] and the latter for for generating random scale-free networks [28], such as the connection of devices on the Internet. Here, we generate two simple cells of seven devices, which are accessed by a central correlator through the nodes which hierarchically have more connectivity (the Power Dominating Set [29], as in [21]), in order to simulate the presence of gateways, as explained in previous sections. The resulting network is depicted in Figure 3, that illustrates the implementation of the Opinion Dynamics correlator and its connection to the rest of nodes, which are labelled in each IIoT cell according to the model used.

Therefore, in this case study we assume the existence of a central correlator that is able to gather the network parameters and the communication information from all devices across each IIoT cell, using the strategies described in Section 3. Afterwards, the virtual agents located in this central correlator will be able to extract features and subsequently form their opinions. In order to show the impact of an attack over their computation, we use the same methodology as in [21]. Here, the authors formalize the attacker model of APTs in a sequence of steps. Each step has its own detection probability, which is quantitatively

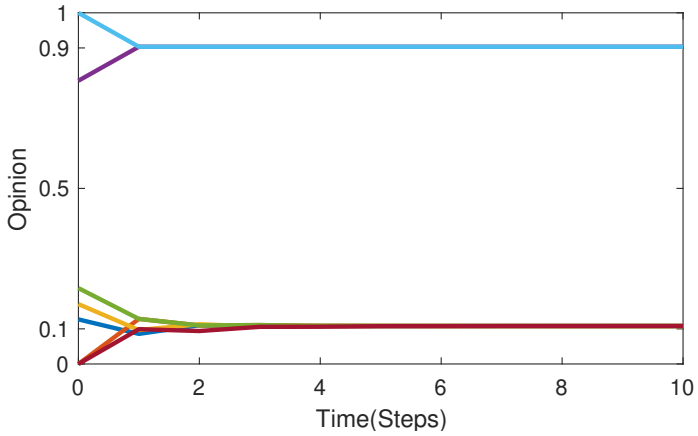


Figure 4: Opinion Dynamics clusters after a lateral movement in the IIoT cell

reflected in each agent to simulate a certain degree of anomaly measured. In our case, we provide these agents with a minimum set of anomaly detection rules based on two features: (1) the delay in their communication channels, and (2) the data values transmitted through those links.

As for the attacker model of this case study, we will base it on the attacker model phases described in [21]. In summary, APTs are sophisticated threats that usually begin with a initial intrusion (e.g., through social engineering or zero-day vulnerabilities), which is followed by stealthy movements throughout the topology until some information is exfiltrated or disruption is caused [14]. Therefore, in this case study we will perpetrate a simple two-step APT attack against the IIoT cell based on the Watts-Strogatz model. These two steps are as follows: an initial intrusion against node 2, and a lateral movement towards node 4. In this basic example, if we consider that this propagation makes use of a covert channel attack (which usually leverages delays introduced arbitrarily in the packet transmissions), then each affected agent should raise a level of anomaly with respect to that feature. This would serve as input to ultimately execute the Opinion Dynamics algorithm and narrow down the attack.

Figure 4 plots the result of the Opinion Dynamics correlation between the seven agents that belong to the Watt-Strogatz cell. The lines represent every agent opinion, that ultimately form two consensus after executing the algorithm with 10 iterations, as explained in Section 2.3. This means that the network is divided into two clusters of nodes that suffer two grades of anomalies: one group of five agents (that sense a 10% of anomaly) and another one of two agents with a 90% that correspond to the nodes involved in the lateral movement. Here, stage S4 has been carried out by *selecting the feature* whose anomaly value is higher, that in the case of node 2 and 3 is the delay. As for the rest of agents, the level of anomaly around 10% appears as consequence of a negligible variation on their data values transmitted. In case that *feature aggregation* was used instead

of selection, an average of the anomalies in both features would be shown on the figure, which only serves as indicator that a greater-than-zero anomaly is occurring. Otherwise, if an *individual Opinion Dynamics instance* were used for each feature, the bottom of the plot in Figure 4 would not appear in the delay one (since those nodes do not show any variation of delay), whereas the top of the plot would not appear in the instance that concerns on the data variation (and opinions of nodes 2 and 3 would also be merged into the bottom cluster due to a low level of variation).

The next step of the framework execution would be to keep track of the multiple APT anomalies over time, associate them with actual attack phases and create a map with the complete threat evolution throughout the network. This is further illustrated with a real setup in [23]. Altogether, this brief example of threat detection exhibits how a security administrator could benefit from different correlation configurations to trace down the implicate nodes of an attack and accurately filter the anomalies suffered across the topology at all levels. This helps to identify the origin of the infection while anticipating further actions to introduce effective response procedures.

5 Conclusions and future work

The degree of sophistication of cyber-security attacks perpetrated against critical infrastructures is increasing world-wide, while the introduction of technologies like the Internet of Things bring benefits but also vulnerabilities across all sectors. Therefore, it is crucial to envision advanced security services beyond traditional measures, being the Opinion Dynamics approach a promising solution that has been proved theoretically in traditional control systems. In this work, we have studied the applicability of this algorithm to the context of the IIoT, providing an insight on how potential instantiations of the Opinion Dynamics algorithm could improve the traceability of attacks in IIoT environments. Our ongoing work includes not only the execution of practical attack cases on a real IIoT testbed to validate these findings, but also additional studies on the advantages and disadvantages of centralized and distributed correlators in this context, plus the potential inclusion of host-based parameters of IIoT devices as data inputs.

Acknowledgments

This work has been supported by the EU H2020-SU-ICT-03-2018 Project No. 830929 CyberSec4Europe (cybersec4europe.eu) and the EU H2020-MSCA-RISE-2017 Project No. 777996 (SealedGRID). Likewise, the work of the first author has been partially financed by the Spanish Ministry of Education under the FPU program (FPU15/03213).

References

[1] Y. Liao, E. de Freitas Rocha Loures, and F. Deschamps, "Industrial Internet of Things: A Systematic Literature Review and Insights," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4515–4525, 2018.

[2] Industrial Internet Consortium, "IIRA Reference Architecture," last accessed 31 August 2019. [Online]. Available: https://www.iiconsortium.org/

[3] Platform Industrie 4.0, "RAMI4.0 Reference Architecture," last accessed 31 August 2019. [Online]. Available: https://www.plattform-i40.de

[4] Siemens, "Industrial IoT Solutions: SIMATIC IoT," last accessed 31 August 2019. [Online]. Available: https://www.siemens.com/

[5] J. E. Rubio, C. Alcaraz, R. Roman, and J. Lopez, "Current cyber-defense trends in industrial control systems," *Computers & Security Journal*, 07/2019 2019.

[6] G. De La Torre, P. Rad, and K.-K. R. Choo, "Implementation of deep packet inspection in smart grids and industrial internet of things: Challenges and opportunities," *Journal of Network and Computer Applications*, 2019.

[7] H. Hadeli, R. Schierholz, M. Braendle, and C. Tuduce, "Leveraging determinism in industrial control systems for advanced anomaly detection and reliable security configuration," in *2009 IEEE Conference on Emerging Technologies & Factory Automation*. IEEE, 2009, pp. 1–8.

[8] L. Zhou and H. Guo, "Anomaly detection methods for iiot networks," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, 2018, pp. 214–219.

[9] R. Berthier and W. H. Sanders, "Specification-based intrusion detection for advanced metering infrastructures," in *IEEE 17th Pacific Rim International Symposium on Dependable Computing*, 2011, pp. 184–193.

[10] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes, "Using model-based intrusion detection for scada networks," in *Proceedings of the SCADA security scientific symposium*, vol. 46. Citeseer, 2007, pp. 1–12.

[11] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389–1407, 2017.

[12] H. Lin, A. Slagell, Z. Kalbarczyk, P. W. Sauer, and R. K. Iyer, "Semantic security analysis of scada networks to detect malicious control commands in power grids," in *Proceedings of the first ACM workshop on Smart energy grid security*. ACM, 2013, pp. 29–34.

[13] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Cyber security of water scada systems—part ii: Attack detection using enhanced hydrodynamic models," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1679–1693, 2012.

[14] A. Lemay, J. Calvet, F. Menet, and J. M. Fernandez, "Survey of publicly available reports on advanced persistent threat actors," *Computers & Security*, vol. 72, pp. 26–59, 2018.

[15] A. Vance, "Flow based analysis of advanced persistent threats detecting targeted attacks in cloud computing," in *2014 First International Scientific-Practical Conference Problems of Infocommunications Science and Technology*. IEEE, 2014, pp. 173–176.

[16] M. Marchetti, F. Pierazzi, M. Colajanni, and A. Guido, "Analysis of high volumes of network traffic for advanced persistent threat detection," *Computer Networks*, vol. 109, pp. 127–141, 2016.

[17] G. Brogi and V. V. T. Tong, "Terminaptor: Highlighting advanced persistent threats through information flow tracking," in *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2016, pp. 1–5.

[18] I. Ghafir, M. Hammoudeh, V. Prenosil, L. Han, R. Hegarty, K. Rabie, and F. J. Aparicio-Navarro, "Detection of advanced persistent threat using machine-learning correlation analysis," *Future Generation Computer Systems*, vol. 89, pp. 349–359, 2018.

[19] C. Alcaraz and J. Lopez, "Wide-area situational awareness for critical infrastructure protection," *Computer*, vol. 46, no. 4, pp. 30–37, 2013.

[20] J. E. Rubio, C. Alcaraz, and J. Lopez, "Preventing advanced persistent threats in complex control networks," vol. 10493. 22nd European Symposium on Research in Computer Security (ESORICS 2017), 2017, pp. 402–418.

[21] J. E. Rubio, R. Roman, C. Alcaraz, and Y. Zhang, "Tracking advanced persistent threats in critical infrastructures through opinion dynamics," in *European Symposium on Research in Computer Security (ESORICS 2018)*, vol. 11098, Springer. Springer, 08/2018 2018, pp. 555–574.

[22] J. E. Rubio, M. Manulis, C. Alcaraz, and J. Lopez, "Enhancing security and dependability of industrial networks with opinion dynamics," in *European Symposium on Research in Computer Security (ESORICS2019)*, vol. 11736, 09/2019 2019, pp. 263–280.

[23] J. E. Rubio, R. Roman, C. Alcaraz, and Y. Zhang, "Tracking apts in industrial ecosystems: A proof of concept," *Journal of Computer Security*, vol. 27, pp. 521–546, 09/2019 2019.

[24] J. Lopez, J. E. Rubio, and C. Alcaraz, "A resilient architecture for the smart grid," *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 3745–3753, 08/2019 2018.

[25] Y. W. Chen, L. F. Zhang, and J. P. Huang, "The watts–strogatz network model developed by including degree distribution: theory and computer simulation," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 29, p. 8237, 2007.

[26] A.-L. Barabási, E. Ravasz, and T. Vicsek, "Deterministic scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 299, no. 3-4, pp. 559–564, 2001.

[27] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the national academy of sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.

[28] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific american*, vol. 288, no. 5, pp. 60–69, 2003.

[29] T. W. Haynes, S. M. Hedetniemi, S. T. Hedetniemi, and M. A. Henning, "Domination in graphs applied to electric power networks," *SIAM Journal on Discrete Mathematics*, vol. 15, no. 4, pp. 519–529, 2002.