

## TRACING WHERE IOT DATA ARE COLLECTED AND AGGREGATED

CHIARA BODEI, PIERPAOLO DEGANO, GIAN-LUIGI FERRARI, AND LETTERIO GALLETTA

Dipartimento di Informatica, Università di Pisa  
*e-mail address:* {chiara,degano,giangi,galletta}@di.unipi.it

**ABSTRACT.** The Internet of Things (IoT) offers the infrastructure of the information society. It hosts smart objects that automatically collect and exchange data of various kinds, directly gathered from sensors or generated by aggregations. Suitable coordination primitives and analysis mechanisms are in order to design and reason about IoT systems, and to intercept the implied technological shifts. We address some of these issues from a foundational point of view. To study them, we define IoT-LySA, a process calculus endowed with a static analysis that tracks the provenance and the manipulation of IoT data, and how they flow in the system. The results of the analysis can be used by a designer to check the behaviour of smart objects, in particular to verify non-functional properties, among which security.

### 1. INTRODUCTION

Nowadays, an increasingly huge number of heterogeneous devices can be easily plugged into a cyber-physical communication infrastructure, the Internet of Things (IoT). “Software is eating the world” is the vivid slogan referring to the *smartification* of the objects and devices around us. The vision offered by IoT as the infrastructure of the information society is fascinating. It amounts to a global network of things, each with a unique identifier, ranging from light bulbs to cars, equipped with suitable software allowing things to interact each other and coordinate their behaviour. Furthermore, smart devices can automatically exchange information of various kinds gathered from different sources (e.g. sensors) or generated by aggregating several data sets.

IoT changes the way we interact with our surroundings. As an example, a smart alarm clock can drive our coffeemaker to prepare us a cup of coffee in the morning; our home automation system turns on our front door light when we arrive at home; then our smart TV can suggest us some movies for the evening, based e.g. on our previous choices.

More connected smart devices and more applications available on the IoT mean more software bugs and vulnerabilities to identify and fix. For instance, a bug can cause you to wake up into a cold house in winter or an attacker can enter into your smart TV or baby monitor and use it to severely deplete service availability, as it seems to be the case with the recent DDoS attacks reported in the news.<sup>1</sup>

---

Partially supported by Università di Pisa PRA\_2016\_64 Project *Through the fog*.

<sup>1</sup><https://www.theguardian.com/technology/2016/oct/22/smart-devices-too-dumb-to-fend-off-cyber-attacks-say-experts>

Smart devices exhibit and require *open-endedness* to achieve full interactive and cooperative behaviour over the Internet. Actually, they generalise the so-called “embedded systems” that essentially are controllers of machines *not* connected to the Internet and therefore we consider them to live in a *closed* world. Consequently, new software solutions have emerged for supporting the design and development of IoT, e.g. Amazon AWS for IoT and Google Brillo. We argue that the standard formal techniques and tools need to be adapted in order to support open-endedness of IoT applications and the new complex phenomena that arise in this hybrid scenario.

Here, we contribute to this emerging line of research by introducing the kernel of a formal design framework for IoT, which will provide us with the foundations to develop verification techniques and tools for checking properties of IoT applications.

Our starting point is the process calculus IoT-LYSA, a dialect of LYSA [8, 16], within the process calculi approach to IoT [32, 34]. It has primitive constructs to model sensors and actuators, and suitable primitives both for processing data and for managing the coordination and communication capabilities of interconnected smart objects. We implicitly assume that sensors are active entities that read data from the physical environment at their own fixed rate. Actuators instead are passive: they just wait for a command to become active and operate on the environment. Briefly, our calculus consists of:

- (1) systems of nodes, made of (a representation of) the physical components, i.e. sensors and actuators, and of software control processes for specifying the *logic* of the objects in the node, including the manipulation of data gathered from sensors and from other nodes. Intra-node generative communications are implemented through a shared store à la Linda [26, 18]. The adoption of this coordination model supports a smooth implementation of the *cyber-physical control architecture*: physical data are made available by sensors to software entities that manipulate them and trigger the relevant actuators to perform the desired actions of the environment.
- (2) a primitive for asynchronous multi-party communication among nodes, which can be easily tuned to take care of various constraints, mainly those concerning proximity;
- (3) functions to process and aggregate data.

Our present version of IoT-LYSA is specifically designed to model monitoring system typical of smart cities, factories or farms. In this scenario, smart objects never leave their locations, while mobile entities, such as cars or people, carry no smart device and can only trigger sensors. For this reason we do not address mobility issues here.

A further contribution of this paper is the definition of an analysis for IoT-LYSA to statically predict the run time behaviour of smart systems. We introduce a Control Flow Analysis that safely approximates the behaviour of a system of nodes. Essentially, it describes the interactions among nodes, tracks how data spread from sensors to the network, and how data are manipulated.

Technically, our analysis abstracts from the concrete values and only considers their provenance and how they are put together and processed, giving rise to *abstract values*. In more detail, it returns for each node  $\ell$  in the network:

- an abstract store  $\widehat{\Sigma}_\ell$  that records for each sensor and each variable a super-set of the abstract values that they may denote at run time;
- a set  $\kappa(\ell)$  that over-approximates the set of the messages received by the node  $\ell$ , and for each of them its sender;
- a set  $\Theta(\ell)$  of possible abstract values computed and used by the node  $\ell$ .

The results of the analysis provide us with the basis for checking and certifying various properties of IoT systems. As it is, the components  $\kappa$  and  $\Theta$  track how data may flow in the network and how they influence the outcome of functions. An example of property that can be statically checked using the component  $\kappa$  is the detection of redundant communications, thus providing the basis for refactoring the system to increase its performance. Another example concerns which nodes, if any, use the values read by a specific sensor, and this property simply requires inspecting the component  $\Theta$  to be established.

In order to assess the applicability of our analysis for verifying IoT systems, we additionally consider some security properties. For that, we extend our core calculus with cryptographic primitives, and propose a general schema in which some classical security policies can be expressed, in particular secrecy and access control. We show then that static checks on the outcome of the analysis help in evaluating the security level of the system and in detecting its potential vulnerabilities.

**Outline of the paper.** The paper is organised as follows. The next section intuitively introduces our proposal with the help of an illustrative example. In Section 3 we introduce the process calculus IOT-LYSA, and we present our Control Flow Analysis in Section 4. We consider security issues in Section 5. Concluding remarks and related work are in Section 6. The appendixes contain all the proofs with some auxiliary results, and a table with the abbreviations and the notation used.

Portions of Sections 2, 3, and 4 appeared in a preliminary form in [11] where (i) sensors did not probe the operating environment and the effects of actuators were not tracked; and (ii) the analysis was less precise, because data were abstracted in a coarser way. Section 5 re-works almost completely our early proposal of checking security policies in [10].

## 2. A SMART STREET LIGHT CONTROL SYSTEM

The IoT European Research Cluster (IERC) has recently identified *smart lighting* in smart cities [30] as one of most relevant applications for the Internet of Things. Recent studies, e.g. [24, 25], show that smart street light control systems represent effective solutions to improve energy efficiency. Many proposed solutions are based on sensors that acquire data about the physical environment and regulate the level of illumination according to the detected events. In this section we show how this kind of scenario can be modelled in IOT-LYSA and what kind of information our Control Flow Analysis provides to designers.

**2.1. System specification.** We consider a simplified system working on a one-way street, inside a restricted traffic zone. It is made of two integrated parts. The first consists of smart lamp posts that are battery powered, can sense their surrounding environment and can communicate with their neighbours to share their views. If (a sensor of) the lamp post perceives a pedestrian and there is not enough light in the street it switches on the light and communicates the presence of the pedestrian to the lamp posts nearby. When a lamp post detects that the level of battery is low, it informs the supervisor of the street lights,  $N_s$ , that will activate other lamp posts nearby. The second component of the street light controller uses the electronic access point to the street. When a car crosses the checkpoint, a message is sent to the supervisor of the street accesses,  $N_a$ , that in turn notifies the presence of the car to  $N_s$ , which acts as a point of connectivity to the Internet. A notice is also sent to the node  $N_{pd}$  that represents a cloud service of the police department. This service checks whether

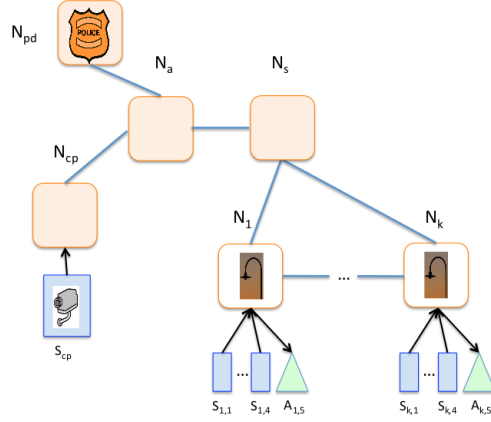


Figure 1: The organisation of nodes in our street light control system.

the car is enabled to enter that restricted zone, through automatic number plate recognition; below we will omit any further detail on this node, e.g. the components for fining the driver. The supervisor  $N_s$  sends a message to the lamp post closest to the checkpoint that starts a forward chain till the end of the street, thus completing the specification of the overall cooperative behaviour. The structure of our control light system is in Figure 1. Below, we will often use a sugared version of the syntax that is made precise in sub-section 3.1.

In our model we assume that each sensor has a unique identifier, hereafter a natural number. Analogously, also actuators have a unique identifier. Since every node has a fixed number of variables, the store of the node can be seen as an array, a portion of which is designated to record the values read by sensors. A sensor identifier is then used as the index to access its reserved store location. To emphasise that indexes are sensor identifiers, we underline them.

In IoT-LYSA each node consists of control processes, sensors and actuators, and of a local store. Processes specify the logic of the system: they coordinate sensors and actuators, communicate with the other nodes and manage data gathered from sensors and from other nodes.

We now define the checkpoint  $N_{cp}$  node. It only contains a visual sensor  $S_{cp}$ , defined below, that takes a picture of the car detected in the street:

$$S_{cp} = \mu h. (\tau. probe(\underline{1})). \tau. h$$

where  $\underline{1}$  is the identifier of the sensor  $S_{cp}$ , and  $probe(\underline{1})$  returns the picture of the car. The sensor makes the picture available to the other components of the node  $N_{cp}$  by storing it in the location (identified by)  $\underline{1}$  of the shared store. The action  $\tau$  denotes internal actions of the sensor, which we are not interested in modelling, e.g. adjusting the camera focus; the construct  $\mu h.$  implements the iterative behaviour of the sensor. Then, the taken picture is enhanced (by using the function  $noiseRed$  for reducing noise) by the process  $P_{cp}$  and sent to the supervisor  $N_a$

$$P_{cp} = \mu h. (z := \underline{1}). (z' := noiseRed(z)). \langle\langle z' \rangle\rangle \triangleright \{\ell_a\}. h$$

where  $\ell_a$  is the label of the node  $N_a$ ; the assignment  $z := \underline{1}$  stores in  $z$  the picture read by the sensor  $S_{cp}$  (recall that  $\underline{1}$  is the identifier of  $S_{cp}$ ); and  $\langle\langle z' \rangle\rangle$  is a multi-output communication

sending  $z'$  to the nodes with labels indicated after  $\triangleright$  (here only to the single node  $N_a$ ). The checkpoint  $N_{cp}$  is defined as

$$N_{cp} = \ell_{cp} : [P_{cp} \parallel S_{cp} \parallel B_{cp}]$$

where  $\ell_{cp}$  is the identifier of  $N_{cp}$  and  $B_{cp}$  abstracts other components we are not interested in, among which its store  $\Sigma_{cp}$ . The node  $N_a$  receives the picture and communicates the presence of the car to the lamp posts supervisor  $N_s$  and to the police department  $N_{pd}$ . The specification of  $N_a$  is as follows

$$N_a = \ell_a : [ \mu h. (; x). \langle\langle car, x \rangle\rangle \triangleright \{ \ell_s, \ell_{pd} \}. h \parallel B_a ]$$

where  $\ell_s$  and  $\ell_{pd}$  are the identifiers of  $N_s$  and  $N_{pd}$ , respectively (see below for the intuition of the general format of the input  $(; x)$ ). The supervisor  $N_s$  contains the process  $P_{s,1}$  that receives the picture from  $N_a$  and sends a message to the node closest to the checkpoint, call it  $N_1$ , labelled with  $\ell_1$ :

$$P_{s,1} = \mu h. (car; x). \langle\langle x \rangle\rangle \triangleright \{ \ell_1 \}. h$$

The input  $(car; x)$  is performed only if the first element of the corresponding output matches the element before the “;” (in this case the constant  $car$ ), and the store variable  $x$  is bound to the value of the second element of the output (see below for the full definition of  $N_s$ ).

In our smart street light control system there is a node  $N_p$  for each lamp post, each of which has a unique identifier  $p \in [1, k]$ . Each lamp post is equipped with four sensors to sense (1) the environment light, (2) the solar light, (3) the battery level and (4) the presence of a pedestrian. We define each of them as follows

$$S_{p,i} = \mu h. probe(i). \tau. h$$

where  $probe(i)$  returns the perceived value by the  $i^{th}$  sensor  $S_{p,i}$  and  $i \in [1, 4]$ , and stores it. After some internal actions  $\tau$ , the sensor  $S_{p,i}$  iterates its behaviour. The actuator for the lamp post  $p$  is defined as

$$A_5 = \mu h. (\underline{5}, \{turnon, turnoff\}). h$$

It only accepts a message from  $N_c$  whose first element is its (undelined) identifier (here  $\underline{5}$ ) and whose second element is either command  $turnon$  or  $turnoff$  and executes it.

The control process of each lamp post node is composed by two parallel processes,  $P_{p,1}$  and  $P_{p,2}$ . The first process is defined as follows

$$\begin{aligned} P_{p,1} = \mu h. & (x_1 := \underline{1}. x_2 := \underline{2}. x_3 := \underline{3}. x_4 := \underline{4}). \\ & (x_4 = true) ? \\ & (x_1 \geq th_1 \wedge x_2 \geq th_2) ? \\ & (x_3 \geq th_3) ? \langle\langle \underline{5}, turnon \rangle\rangle. \langle\langle x_4 \rangle\rangle \triangleright L_p. h \\ & : \langle\langle err, \ell_p \rangle\rangle \triangleright \{ \ell_s \}. h \\ & : h \\ & : \langle\langle \underline{5}, turnoff \rangle\rangle. h \end{aligned}$$

The process reads the current values from the sensors and stores them into the local variables  $x_i$ . The nested conditional statement says that the actuator is turned on if (i) a pedestrian is detected in the street ( $x_4$  holds), (ii) the intensity of environment and solar lights are greater than or equal to the given thresholds  $th_1$  and  $th_2$ , and (iii) there is enough battery (at least  $th_3$ ). In addition, the presence of the pedestrian is communicated to the lamp

posts nearby, whose labels, typically  $\ell_{p-1}$  and  $\ell_{p+1}$ , are in  $L_p$ . Instead, if the level battery is insufficient, an error message, including its identifier  $\ell_p$ , is sent to the supervisor node, labelled  $\ell_s$ . The second process  $P_{p,2}$  is defined as follows

$$P_{p,2} = \mu h.(\cdot; x). (x = true \vee is\_a\_car(x)) ? (\langle \bar{5}, turnon \rangle. \langle \langle x \rangle \rangle \triangleright L_p). h : \langle \bar{5}, turnoff \rangle. h$$

It waits for messages from its neighbours or from the supervisor node  $N_s$ . When one of them is notified the presence of a pedestrian ( $x = true$ ) or of a car ( $is\_a\_car(x)$  holds), the current lamp post orders the actuator to switch the light on. Each lamp post  $p$  is described as the IOT-LYSA node below

$$N_p = \ell_p : [\Sigma_p \parallel P_{p,1} \parallel P_{p,2} \parallel S_{p,1} \parallel S_{p,2} \parallel S_{p,3} \parallel S_{p,4} \parallel A_{p,5}]$$

where  $\Sigma_p$  is the store of the node  $\ell_p$ , shared among its components. The supervisor node  $N_s$  of lamp posts is defined below, where  $x$  ranges over the  $k$  lamp posts

$$N_s = \ell_s : [\mu h. (err; x). \langle \langle true \rangle \rangle \triangleright L_x. h \parallel P_{s,1} \parallel B_s]$$

where  $P_{s,1}$  is the process previously defined. As above the input  $(err; x)$  is performed only if the first element of the corresponding output matches the constant  $err$ , and the store variable  $x$  is bound to the value of the second element of the output, i.e. the label of the relevant lamp post. If this is the case, after some internal computations,  $N_s$  warns the lamp posts nearby  $x$  (included in  $L_x$ ) of the presence of a pedestrian.

The whole intelligent controller  $N$  of the street lights is then described as the parallel composition of the checkpoint node  $N_{cp}$ , the supervisors nodes  $N_a$  and  $N_s$ , the nodes of lamp posts  $N_p$ , with  $p \in [1, k]$ , and the police department node  $N_{pd}$ :

$$N = N_{cp} \mid N_a \mid N_s \mid N_1 \mid \dots \mid N_k \mid N_{pd}$$

**2.2. Checking properties.** We would like to statically predict how the system behaves at run time. In particular, we want to compute: (i) how nodes interact each other; (ii) how data spread from sensors to the network (tracking); and (iii) which computations each node performs on the received data. To do that, we define a Control Flow Analysis, which abstracts from the concrete values by only considering their provenance and how they are manipulated. For example, consider the picture sent by the camera of  $S_{cp}$  to its control process  $P_{pc}$ . In the analysis we are only interested in tracking where the picture comes from, and not in its actual value; so we use (a suitable representation of) the abstract value  $1^{\ell_{cp}}$  to record the camera that took it. The process  $P_{pc}$  reduces the noise in the pictures and sends the result to  $N_a$ . Our analysis keeps track of this manipulation through (a representation of) the abstract value  $noiseRed^{\ell_{cp}}(1^{\ell_{cp}})$ , meaning that the function  $noiseRed$ , computed by the node  $\ell_{cp}$ , is applied to data coming from the sensor with identifier 1 of  $\ell_{cp}$ .

In more detail, our analysis returns for each node  $\ell$  in the network: an abstract store  $\widehat{\Sigma}_\ell$  that records for each variable a super-set of the abstract values that it may denote at run time and for each sensor a specific abstract value as exemplified above; a set  $\kappa(\ell)$  that approximates the set of the messages received by the node  $\ell$ ; and the set  $\Theta(\ell)$  of possible abstract values computed and used by the node  $\ell$ . Actually, abstract values may grow unbounded and we will thus represent them through regular tree grammars. In our example, for each lamp post labelled  $\ell_p$ , the analysis returns in  $\kappa(\ell_p)$  both (the grammar representing) the abstract value  $noiseRed^{\ell_{cp}}(1^{\ell_{cp}})$  and the sender of that message, i.e.  $\ell_{p+1}$ .

The result of our analysis can be exploited to perform several verifications. For instance, since the pictures of cars are sensitive data, one would like to check whether they are kept confidential. A simple check on  $\kappa$  suffices for discovering that our system fails to protect the pictures because they are always sent in clear. An obvious solution is to encrypt these sensitive data before sending them. However, privacy is not guaranteed either, because the encrypted information is correctly sent to the supervisor of the street access  $N_a$  and to the police department  $N_{pd}$ , but also to all lamp posts. By inspecting  $\kappa$  and  $\Theta$  we detect this violation. A possible solution is making the picture anonymous through a suitable function  $an$ , e.g. blurring the plate, before sending it to the supervisor of the street light  $N_s$ . In order to reach a balance between protecting data for privacy and using them for safety, we can resort to the following amended code of the process  $P_{cp}$  and of the node  $N_a$ :

$$P'_{cp} = \mu h.(z := \underline{1}).(z' := \text{noiseRed}(z)).\langle\langle\{z'\}_k\rangle\rangle \triangleright \{\ell_a\}. h$$

$$N'_a = \ell'_a : [\mu h.(; y).\text{decrypt } y \text{ as } \{; x\}_k \text{ in } \langle\langle car, \{x\}_{k'}\rangle\rangle \triangleright \{\ell_{pd}\}. \langle\langle car, an(x)\rangle\rangle \triangleright \{\ell_s\}. h \parallel B_a]$$

Another way of exploiting the results of our analysis is detecting whether there are redundant communications, which are possibly power consuming. For example, since the street is one-way, when a car is present the lamp post at position  $p$  needs not to alert the one at  $p - 1$ . By inspecting  $\kappa$  it is easy to detect a redundant, useless communication from the next lamp post. On this basis, the designer can remove the label  $\ell_{p-1}$  from the set  $L_p$  of receivers in the definition of  $P_{p,2}$  for all lamp posts.

A further issue concerns the fact that an IoT system should be robust enough and work in presence of a faulty device. Our analysis can support a “what-if” reasoning. For example, suppose that the designer wants to check what happens if a lamp post, say  $\ell_{p-1}$ , goes out of order. By inspecting the  $\kappa$  component, the designer discovers that  $\ell_p$  may receive messages from  $\ell_{p-1}$  and  $\ell_{p+1}$ . Assuming that redundancy has been removed as sketched above, no message will be sent to  $\ell_p$  from  $\ell_{p-1}$  when a car is in the street (as specified in the process  $P_{i-1,2}$ ). Consequently, all the lamp posts  $\ell_q$  with  $q \geq p - 1$ , will not be switched on, even though a car is running in the street. Clearly, when pedestrians are walking there, the relevant lamp post will be switched on as soon as their presence is detected, except for the faulty one, of course. An easy fix is enlarging the set  $L_p$  in the process  $P_{p,2}$  to also contain  $\ell_{p+2}$ , so introducing a possibly useful redundancy; similarly, the process  $P_{s,1}$  will send its message to both  $\ell_1$  and  $\ell_2$ .

### 3. THE CALCULUS IOT-LYSA

The IoT applications operate in a cyber-physical world, and therefore modelling them requires taking into account both logical and physical aspects. Physical data, typically collected by environmental sensors, influence the logics of an application, which in turn modifies the physical world, through actuators. Here we are only interested in modelling the logical components, in identifying their frontier with the physical world and in abstractly representing the interactions between them. Typical representations of the world are based on continuous models, like, e.g. (ordinary or stochastic) differential equations, or on discrete version of them or on hybrid automata [28]. Instead, for us the physical world is a black box from which sensors, that are always active, can measure the value of some observables with a certain rate of their own. These values are made available to controllers that trigger suitable actuators. Actuators are passive entities that can only execute their task when activated; they operate on the physical world, so changing the value of some observables.

Furthermore, we assume that the world can autonomously evolve, and that the changes of its state are revealed and detected by sensors.

In order to model the logical components of IoT applications, we adapt the LYSA calculus [7, 8, 16], based on the  $\pi$ - [36] and the Spi-calculus [1]. For that we introduce:

- (i) systems of nodes, in turn consisting of sensors, actuators and control processes, plus a shared store within each node, allowing for internal communications;
- (ii) primitives for measuring values in the world with sensors, and for triggering actuator actions;
- (iii) an asynchronous multi-party communication modality among nodes, subject to constraints, mainly concerning physical proximity;
- (iv) functions to process data;
- (v) explicit conditional statements.

We also extend our proposal in Section 5 with

- (vi) encryption and decryption constructs to represent and handle some logical aspects of security.

**3.1. Syntax.** The logical components of an IoT system are specified using a two-level syntax, one describing the whole system and the other its components. At the first level one defines system, consisting of the parallel composition of a fixed number of labelled nodes. Each node is defined at the second level, and it hosts a store, control processes, sensors and actuators. The label  $\ell$  uniquely identifies the node  $\ell : [B]$  and may represent further characterising information (e.g. its location or other contextual information). The syntax of systems is as follows.

$\mathcal{N} \ni N ::=$	<i>systems of nodes</i>	
	$0$	empty system
	$\ell : [B]$	single node ( $\ell \in \mathcal{L}$ , the set of labels)
	$N_1 \mid N_2$	parallel composition of nodes
$\mathcal{B} \ni B ::=$	<i>node components</i>	
	$\Sigma_\ell$	store of node $\ell$
	$P$	process
	$S$	sensor, with a unique identifier $i \in \mathcal{I}_\ell$
	$A$	actuator, with a unique identifier $j \in \mathcal{J}_\ell$
	$B \parallel B$	parallel composition of node components

A node component  $B$  contains  $\#(\mathcal{I}_\ell)$  sensors  $S$  and  $\#(\mathcal{J}_\ell)$  actuators  $A$ . It also has finitely many control processes  $P$  that use a finite set of variables  $\mathcal{X}_\ell$ . We impose that in  $\ell : [B]$  there always is a *single* store  $\Sigma_\ell : \mathcal{X}_\ell \cup \mathcal{I}_\ell \rightarrow \mathcal{V}$ , where  $\mathcal{V}$  is the denumerable set of values, including numbers, booleans etc., but neither labels nor identifiers of sensors and actuators. (We feel free to omit from here afterwards the label  $\ell$  when immaterial.) Therefore, a store is essentially an array of fixed dimension, and intuitively a variable  $x \in \mathcal{X}_\ell$  and an identifier  $i \in \mathcal{I}_\ell$  are interpreted as indexes in the array (no need of  $\alpha$ -conversions). We assume that store accesses are atomic, e.g. through CAS instructions [29].



The syntax of control processes is as follows

$P ::=$	<i>control processes</i>	
	$0$	inactive process
	$\langle\langle E_1, \dots, E_r \rangle\rangle \triangleright L.P$	asynchronous multi-output $L \subseteq \mathcal{L}$
	$(E_1, \dots, E_j; x_{j+1}, \dots, x_r).P$	input (with matching)
	$E?P : Q$	conditional statement
	$h$	iteration variable
	$\mu h. P$	iteration
	$x := E.P$	assignment to $x \in \mathcal{X}_\ell$
	$\langle j, \gamma \rangle.P$	output of action $\gamma$ to actuator $j$

The prefix  $\langle\langle E_1, \dots, E_r \rangle\rangle \triangleright L$  implements a simple form of multicast communication among nodes: the tuple  $E_1, \dots, E_r$  is asynchronously sent to the nodes with labels in  $L$  and that are “compatible” (according, among other attributes, to a proximity-based notion and to the transmission capability of the sender). The input prefix  $(E_1, \dots, E_j; x_{j+1}, \dots, x_r)$  is willing to receive a  $r$ -tuple, provided that its first  $j$  terms match the input ones, and then binds the remaining store variables (separated by a “;”) to the corresponding values (see [16, 6] for a more flexible choice). Otherwise, the  $r$ -tuple is not accepted. A process repeats its behaviour, when defined through the iteration construct  $\mu h. P$ , where  $h$  is the iteration variable; an obvious sanity requirement is that any recursion variable  $h$  only occurs within the scope of its binding  $\mu h$ . Finally, a process can command an actuator to perform an action over the physical world.

Sensors and actuators have the form:

$S ::=$	<i>sensors</i>	$A ::=$	<i>actuators</i>
	$0$	$0$	inactive actuator
	$\tau.S$	$\tau.A$	internal action
	$probe(i).S$	$\langle j, \Gamma \rangle.A$	command for actuator $j$
		$\gamma.A$	triggered action ( $\gamma \in \Gamma$ )
	$h$	$h$	iteration variable
	$\mu h.S$	$\mu h.A$	iteration

We recall that each sensor and each actuator is identified by a unique identifier belonging to the sets  $\mathcal{I}_\ell$  and  $\mathcal{J}_\ell$ , respectively. A sensor can perform an internal action  $\tau$ , e.g. for resetting or for changing its battery mode. In addition, it senses the physical world and stores the value observed in its location  $i$ . Recall that each sensor is dedicated to measure a specific observable, e.g. the temperature or the humidity. Also an actuator can perform an internal action  $\tau$ . More interestingly, upon receiving a command  $\gamma$  from a controller, an actuator executes  $\gamma$ , possibly causing a change in the physical state of the world. Both sensors and actuators can iterate their behaviour.

The syntax of terms follows.

$E ::=$	<i>terms</i>	
	$v$	value ( $v \in \mathcal{V}$ )
	$i$	sensor location ( $i \in \mathcal{I}_\ell$ )
	$x$	variable ( $x \in \mathcal{X}_\ell$ )
	$f(E_1, \dots, E_r)$	function on data ( $f \in \mathcal{F}$ )

The term  $f(E_1, \dots, E_r)$  is the application of function  $f$  to  $n$  arguments; we assume as given a set of primitive functions  $\mathcal{F}$ , typically for aggregating or comparing values, be they computed or data sampled from the environment.

**3.2. Operational Semantics.** Our reduction semantics relies on an abstract model of the evolution of the physical world from a state  $\mathcal{E}$  to a state  $\mathcal{E}'$ . We write  $\mathcal{E} \triangleright \mathcal{E}'$  when a state transition occurs, without detailing how this happens, because for us the world is a black box. Similarly, we will write  $\gamma(\mathcal{E}) \triangleright \mathcal{E}'$  to represent the state transition caused by an actuator that performs the action  $\gamma$  on  $\mathcal{E}$ .

We take the syntactic elements of IOT-LYSA up to the following *structural congruence*  $\equiv$  on nodes, processes, sensors and actuators. It is standard except for the last rule that equates a multi-output with no receivers and the inactive process, and for the fact that inactive components of a node are all coalesced.

- $(\mathcal{N}/\equiv, |, 0)$  and  $(\mathcal{B}/\equiv, \parallel, 0)$  are commutative monoids
- $\mu h . X \equiv X \{ \mu h . X / h \}$  for  $X \in \{P, A, S\}$  (★)
- $\langle\langle E_1, \dots, E_r \rangle\rangle : \emptyset. 0 \equiv 0$

We have a two-level *reduction relation*  $\xrightarrow[\mathcal{E}']{\mathcal{E}}$  reflecting the two-level structure of IOT-LYSA. It is defined as the least relation on both nodes and their components, satisfying the set of inference rules in Table 1. The arrow is decorated with  $\mathcal{E}$ , the state of the world before the transition, and  $\mathcal{E}'$  after its occurrence. We assume the standard denotational interpretation  $\llbracket E \rrbracket_{\Sigma}$  for evaluating terms.

The first two rules implement the (atomic) asynchronous update of shared variables inside nodes, by using the standard notation  $\Sigma\{-/-\}$  for store update. According to (Sense), the  $i^{\text{th}}$  sensor measures the value  $v$  of its observable through the semantic function *measure* that operates on the environment; then it stores  $v$  into the location  $i$ . The rule (Asgm) specifies how a control process updates the variable  $x$  with the value of  $E$ .

The rules (Ev-out) and (Multi-com) drive asynchronous multi-communications among nodes. In the first a node sends a tuple of values  $\langle\langle v_1, \dots, v_r \rangle\rangle$ , obtained by the evaluation of  $\langle\langle E_1, \dots, E_r \rangle\rangle$ . Asynchrony is realised by spawning the new process  $\langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L.0$  (with the inactive process as continuation) in parallel with the continuation  $P$ ; the new process offers the message to all the receivers belonging to the set  $L$ . In the rule (Multi-com), the message coming from  $\ell_1$  is received by a node labelled  $\ell_2$ . The communication succeeds, provided that (i)  $\ell_2$  belongs to the set  $L$  of possible receivers, (ii) the receiver is within the transmission range of the sender, according to the compatibility function *Comp*, and (iii) that the first  $j$  values match the evaluations of the first  $j$  terms in the input. Moreover, the label  $\ell_2$  is removed by the set of receivers  $L$  of the message. The spawned process terminates when all its receivers have received the message (see the last congruence rule in ★). The role of the compatibility function *Comp* is crucial in modelling real world constraints on communication. A common requirement is that inter-node communications are proximity-based, i.e. that only nodes that are in the transmission range of the sender can read the message. This is easily encoded here by defining a possibly non-symmetric predicate (over node labels) yielding true only when the second node is in the transmission range of the first. Of course, this function could be enriched in order to consider finer notions of compatibility expressing various policies, e.g. topics for event notification.

<p>(Sense)</p> $\frac{measure(i, \mathcal{E}) = v}{\Sigma \parallel probe(i). S \parallel B \xrightarrow[\mathcal{E}]{\varepsilon} \Sigma\{v/i\} \parallel S \parallel B}$	<p>(Asgm)</p> $\frac{\llbracket E \rrbracket_{\Sigma} = v}{\Sigma \parallel x := E. P \parallel B \xrightarrow[\mathcal{E}]{\varepsilon} \Sigma\{v/x\} \parallel P \parallel B}$		
<p>(Ev-out)</p> $\frac{\bigwedge_{i=1}^r \llbracket E_i \rrbracket_{\Sigma} = v_i}{\Sigma \parallel \langle\langle E_1, \dots, E_r \rangle\rangle \triangleright L. P \parallel B \xrightarrow[\mathcal{E}]{\varepsilon} \Sigma \parallel \langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L. 0 \parallel P \parallel B}$			
<p>(Multi-com)</p> $\frac{\ell_2 \in L \wedge Comp(\ell_1, \ell_2) \wedge \bigwedge_{i=1}^j \llbracket E_i \rrbracket_{\Sigma_2} = v_i}{\ell_1 : [\langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L. 0 \parallel B_1] \mid \ell_2 : [\Sigma_2 \parallel (E_1, \dots, E_j; x_{j+1}, \dots, x_r). Q \parallel B_2] \xrightarrow[\mathcal{E}]{\varepsilon} \ell_1 : [\langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L \setminus \{\ell_2\}. 0 \parallel B_1] \mid \ell_2 : [\Sigma_2\{v_{j+1}/x_{j+1}, \dots, v_r/x_r\} \parallel Q \parallel B_2]}$			
<p>(Cond)</p> $\frac{\llbracket E \rrbracket_{\Sigma} = \mathbf{b}_i \quad i \in \{1, 2\}}{\Sigma \parallel E? P_1 : P_2 \parallel B \xrightarrow[\mathcal{E}]{\varepsilon} \Sigma \parallel P_i \parallel B} \quad \text{with } \mathbf{b}_1 = \mathbf{true}, \mathbf{b}_2 = \mathbf{false}$	<p>(Int)</p> $\frac{}{\tau. X \xrightarrow[\mathcal{E}]{\varepsilon} X}$		
<p>(A-com)</p> $\frac{\gamma \in \Gamma}{\langle j, \gamma \rangle. P \parallel (j, \Gamma). A \parallel B \xrightarrow[\mathcal{E}]{\varepsilon} P \parallel \gamma. A \parallel B}$	<p>(Act)</p> $\frac{\gamma(\mathcal{E}) \triangleright \mathcal{E}'}{\gamma. A \xrightarrow[\mathcal{E}']{\varepsilon} A}$	<p>(Phys)</p> $\frac{\mathcal{E} \triangleright \mathcal{E}'}{N \xrightarrow[\mathcal{E}']{\varepsilon} N}$	
<p>(Node)</p> $\frac{B \xrightarrow[\mathcal{E}']{\varepsilon} B'}{\ell : [B] \xrightarrow[\mathcal{E}']{\varepsilon} \ell : [B']}$	<p>(ParN)</p> $\frac{N_1 \xrightarrow[\mathcal{E}']{\varepsilon} N'_1}{N_1 \mid N_2 \xrightarrow[\mathcal{E}']{\varepsilon} N'_1 \mid N_2}$	<p>(ParB)</p> $\frac{B_1 \xrightarrow[\mathcal{E}']{\varepsilon} B'_1}{B_1 \parallel B_2 \xrightarrow[\mathcal{E}']{\varepsilon} B'_1 \parallel B_2}$	<p>(CongrY)</p> $\frac{Y'_1 \equiv Y_1 \xrightarrow[\mathcal{E}']{\varepsilon} Y_2 \equiv Y'_2}{Y'_1 \xrightarrow[\mathcal{E}']{\varepsilon} Y'_2}$

Table 1: Reduction semantics, where  $X \in \{S, A\}$ , and  $Y \in \{N, B\}$ .

According to the evaluation of the expression  $E$ , the rule (Cond) says that the process continues as  $P_1$  (if  $\llbracket E \rrbracket_{\Sigma}$  is true) or as  $P_2$  (otherwise). The rule (Int) applies to sensors and actuators and simply governs the occurrence of internal actions.

A process commands the  $j^{\text{th}}$  actuator through the rule (A-com), by sending it the pair  $\langle j, \gamma \rangle$ . The effect is that  $\gamma$  prefixes the actuator, if  $\gamma$  is one of its actions. The rule (Act) says that the actuator performs the action  $\gamma$  over the current state  $\mathcal{E}$  of the environment, triggering an evolution to a new state. The other rule that changes the environment state is (Phys). It models the fact that environments can evolve independently of the application in hand. Since for us an environment is a black box, we resort to non-determinism to abstract from this kind of independence.

The rules (ParN) and (ParB) propagate reductions across parallel composition; the rule (Node) lifts the transitions from components to nodes. Finally, the rules (CongrY), with  $Y \in \{N, B\}$ , for nodes and components are the standard reduction rules for the congruence defined in ( $\star$ ).

Back to our example of Section 2, consider a run where a picture of a car is taken by the camera  $S_{cp}$  in the node  $\ell_{cp}$  and is sent to the node  $\ell_a$ . First of all, we briefly recall the definition of the camera  $S_{cp}$  and of the control process  $P_{cp}$  of  $\ell_{cp}$ :

$$\begin{aligned} S_{cp} &= \mu h.(\tau.probe(\underline{1})).\tau. h \\ P_{cp} &= \mu h.(z := \underline{1}). \underbrace{(z' := noiseRed(z)).\langle\langle z' \rangle\rangle \triangleright \{\ell_a\}. h}_{P'_{cp}} \end{aligned}$$

The sequence of transitions is as follows:

$$N_{cp} = \ell_{cp}[P_{cp} \parallel S_{cp} \parallel \Sigma_{cp} \parallel B_{cp}] \xrightarrow[\mathcal{E}]{\mathcal{E}^+} \quad (3.1)$$

$$\ell_{cp}[P_{cp} \parallel \underbrace{\tau.S_{cp}}_{S'_{cp}} \parallel \underbrace{\Sigma_{cp}\{v/1\}}_{\Sigma^1_{cp}} \parallel B_{cp}] \xrightarrow[\mathcal{E}]{\mathcal{E}^+} \quad (3.2)$$

$$\ell_{cp}[P'_{cp} \parallel S'_{cp} \parallel \underbrace{\Sigma^1_{cp}\{v/z\}}_{\Sigma^2_{cp}} \parallel B_{cp}] \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.3)$$

$$\ell_{cp}[\langle\langle z' \rangle\rangle \triangleright \{\ell_a\}. P_{cp} \parallel S'_{cp} \parallel \underbrace{\Sigma^2_{cp}\{v'/z'\}}_{\Sigma^3_{cp}} \parallel B_{cp}] \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.4)$$

$$\ell_{cp}[\underbrace{P_{cp} \parallel S'_{cp} \parallel \Sigma^3_{cp} \parallel B_{cp}}_{B'_{cp}} \parallel \langle\langle v' \rangle\rangle \triangleright \{\ell_a\}. 0] = N'_{cp} \quad (3.5)$$

where as usual  $\xrightarrow[\mathcal{E}]{\mathcal{E}^+}$  stands for the occurrence of one or more transitions. The rules (CongrB), (Int) and (Sense) are applied to move from (3.1) to (3.2) and the value  $v$  amounts to  $measure(1, \mathcal{E})$ ; the rules (CongrB) and (Asgm) drive the transition from (3.2) to (3.3); once again the rule (Asgm) drives the transition from (3.3) to (3.4) where the value  $v'$  is the result of applying the function  $noiseRed$  to the value  $v$ ; in the transition from (3.4) to (3.5) the rule (Ev-out) is applied.

As a further example, consider the steps carried out by the node  $\ell_a$  to receive the car picture and to forward it to the policy station  $\ell_{pd}$  and to the lamp post supervisor  $\ell_s$ . Here we recall the definition of the node  $N_a$ :

$$N_a = \ell_a[ \underbrace{\mu h.(; x).\langle\langle car, x \rangle\rangle \triangleright \{\ell_s, \ell_{pd}\}. h}_{P_a} \parallel \Sigma_a \parallel B_a ]$$

The transitions are as follows:

$$N'_{cp} \mid N_a \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.6)$$

$$\ell_{cp}[B'_{cp} \parallel \langle\langle v' \rangle\rangle \triangleright \{\ell_a\}. 0] \mid \ell_a[ (; x).\langle\langle car, x \rangle\rangle \triangleright \{\ell_s, \ell_{pd}\}. P_a \parallel \Sigma_a \parallel B_a ] \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.7)$$

$$\ell_{cp}[B'_{cp} \parallel \langle\langle v' \rangle\rangle \triangleright \emptyset. 0] \mid \ell_a[ \langle\langle car, x \rangle\rangle \triangleright \{\ell_s, \ell_{pd}\}. P_a \parallel \underbrace{\Sigma_a\{v'/x\}}_{\Sigma'_a} \parallel B_a ] \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.8)$$

$$\ell_{cp}[B'_{cp}] \mid \ell_a[ \langle\langle car, x \rangle\rangle \triangleright \{\ell_s, \ell_{pd}\}. P_a \parallel \underbrace{\Sigma_a\{v'/x\}}_{\Sigma'_a} \parallel B_a ] \xrightarrow[\mathcal{E}]{\mathcal{E}} \quad (3.9)$$

$$\ell_{cp}[B'_{cp}] \mid \ell_a[ P_a \parallel \Sigma'_a \parallel B_a \parallel \langle\langle car, v' \rangle\rangle \triangleright \{\ell_s, \ell_{pd}\}.0 ] \quad (3.10)$$

where we apply the rule (CongrB) to the process  $P_a$  for the step from (3.6) to (3.7); the rule (Multi-com) is used to move from (3.7) to (3.8), where  $v'$  is the value sent by the node  $\ell_a$ ; since in the message  $\langle\langle v' \rangle\rangle \triangleright \emptyset$ . 0 the receiver set is empty, the rule (CongrB) drives the transition from (3.8) to (3.9); the last transition from (3.9) to (3.10) is performed by applying the rule (Ev-out) where  $v'$  is forwarded to nodes  $\ell_s$  and  $\ell_{pd}$ . Of course the transitions of part of the nodes are preserved by the whole system of nodes  $N$ , by applying the (ParN) rule.

#### 4. CONTROL FLOW ANALYSIS

This section introduces a static analysis for IOT-LYSA that safely approximates the abstract behaviour of a system of nodes  $N$ , regardless of the environment that hosts it. Our analysis tracks the usage of sensor values inside the local node where they are gathered and their propagation in the network of nodes both as raw data or processed via suitable functions. It also describes which messages a node can receive and from which nodes, so abstractly representing the communication structure of the system.

Technically, we define a Control Flow Analysis (CFA, for short), specified in terms of Flow Logic [40], a declarative approach borrowing from and integrating many classical static techniques [20, 27, 41, 31]. The distinctive feature of Flow Logic is to separate the *specification* of the analysis from its actual *computation*. Intuitively, the specification describes when its results, namely analysis *estimates*, are *valid*. Formally, a specification consists of a set of clauses defining the validity of estimates. Furthermore, Flow Logic provides us with a methodology to define a correct and efficient analysis algorithm, by reducing the specification to a constraint satisfaction problem.

Below, we specify our analysis in a logical form by inducing on the syntax of the constructs of IOT-LYSA along the line of [8]. An algorithmic version of our analysis can be easily derived from its logical specification. It suffices to induce on the syntax for generating a set of constraints in AFPL, a logic used to specify static analyses [39]. The generated constraints are solved in low polynomial time obtaining the minimal valid estimate through the succinct solver presented in [39].

**4.1. Abstract representation of data.** In the following we represent data generated by an IoT system as trees. The leaves of such trees either identify the sensor from which an observable comes or represent a basic value, and the nodes represent the aggregation functions applied to data. Since a system is designed to be continuously active and may contain feedback loops, aggregated data can grow unbounded, and so are the trees used to abstractly represent them. In order to have a finite representation we shall rely here on regular tree grammars [19], rather than coalescing all trees exceeding a given depth in a special abstract value as in [11, 10]. These new abstractions better approximate actual data, and the results of the analysis are accordingly more precise.

A *regular tree grammar* is a quadruple  $\widehat{G} = (\mathbb{N}, \mathbb{T}, Z, R)$  where

- $\mathbb{N}$  is a set of non-terminals (with rank 0),
- $\mathbb{T}$  is a ranked alphabet, whose symbols have an associated arity,
- $Z \in \mathbb{N}$  is the starting non-terminal,
- $R$  is a set of productions of the form  $A \rightarrow t$ , where  $t$  is a tree composed from symbols in  $\mathbb{N} \cup \mathbb{T}$  according to their arities.

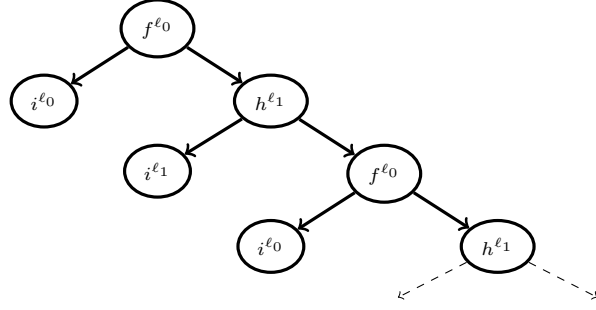


Figure 2: An infinite abstract tree

In the following we denote the language generated by a given grammar  $\widehat{G}$  with  $\text{Lang}(\widehat{G})$ .

Given a system of nodes  $N$ , the grammars we use will have the alphabet  $\mathbb{T}$  consisting of the following set of ranked symbols

- $i^\ell$  (with arity 0) for each sensor  $i \in \mathcal{I}_\ell$
- $v^\ell$  (with arity 0) for each node  $\ell \in \mathcal{L}$
- $f^\ell$  (with arity  $r$ ) for each function  $f \in \mathcal{F}$  and  $\ell \in \mathcal{L}$

The non-terminals  $\mathbb{N}$  of our grammars include a symbol for each terminal, and just for readability we shall use their capital counterparts, i.e.  $I^\ell, V^\ell$  and  $F^\ell$ . In this way, the production  $F^\ell \rightarrow f^\ell(t_1, \dots, t_r)$  generates the tree rooted in  $f^\ell$  and children generated by  $t_1, \dots, t_r$ .

It is convenient introducing some notation. For brevity and when no ambiguity may arise, we will simply write  $\hat{v} = (Z, R)$  for the grammar  $\widehat{G} = (\mathbb{N}, \mathbb{T}, Z, R)$  with starting non-terminal  $Z$  and regular productions in  $R$ , without explicitly listing the terminals and the non-terminals. Then, we denote with  $\mathbb{R}$  the set of all possible productions over  $\mathbb{N}$  and  $\mathbb{T}$ .

As an example of a possible infinite abstract tree, consider two nodes  $N_{\ell_0}$  and  $N_{\ell_1}$  with two aggregation functions  $f$  and  $h$ . Suppose that  $N_{\ell_0}$  applies  $f$  to a value read by the sensor  $i_0$  and a value received from  $N_{\ell_1}$ . Similarly,  $N_{\ell_1}$  applies  $h$  to a value read by the sensor  $i_1$  and the value received from  $N_{\ell_0}$ . The resulting value is abstracted as the possibly infinite binary tree in Figure 2 that belongs to the language of the following grammar:

$$(F^{\ell_0}, \{F^{\ell_0} \rightarrow f^{\ell_0}(I_0^{\ell_0}, H^{\ell_1}), I_0^{\ell_0} \rightarrow i_0^{\ell_0}, H^{\ell_1} \rightarrow h^{\ell_1}(I_1^{\ell_1}, F^{\ell_0}), I_1^{\ell_1} \rightarrow i_1^{\ell_1}\})$$

Now we are ready to introduce the *abstract terms* that belong to the set

$$\widehat{\mathcal{V}} = 2^{\mathbb{N} \times \mathbb{R}}$$

**4.2. Specification of the analysis.** The result of our CFA is a pair  $(\widehat{\Sigma}, \Theta)$  for terms  $E$  and a triple  $(\widehat{\Sigma}, \kappa, \Theta)$  for  $N$ , called *estimate* for  $E$  and for  $N$ , respectively. The components of an estimate are the following *abstract domains*:

- *abstract store*  $\widehat{\Sigma} = \bigcup_{\ell \in \mathcal{L}} \widehat{\Sigma}_\ell : \mathcal{X} \cup \mathcal{I}_\ell \rightarrow 2^{\widehat{\mathcal{V}}}$  where each *abstract local store*  $\widehat{\Sigma}_\ell$  approximates the concrete local store  $\Sigma_\ell$ . It associates with each location a set of abstract values representing the possible concrete values that the location may store at run time.
- *abstract network environment*  $\kappa : \mathcal{L} \rightarrow \mathcal{L} \times \bigcup_{i=1}^m \widehat{\mathcal{V}}^i$  where  $m$  is the maximum arity of the messages exchanged in the system under analysis (with  $\widehat{\mathcal{V}}^{i+1} = \widehat{\mathcal{V}} \times \widehat{\mathcal{V}}^i$ ). Intuitively,  $\kappa$  includes all the abstract messages that may be received by the node labelled  $\ell$ .

$$\begin{array}{c}
\text{(E-sen)} \quad \frac{(I^\ell, \{I^\ell \rightarrow i^\ell\}) \in \vartheta \subseteq \Theta(\ell)}{(\widehat{\Sigma}, \Theta) \models_\ell i : \vartheta} \qquad \text{(E-val)} \quad \frac{(V^\ell, \{V^\ell \rightarrow v^\ell\}) \in \vartheta \subseteq \Theta(\ell)}{(\widehat{\Sigma}, \Theta) \models_\ell v : \vartheta} \qquad \text{(E-var)} \quad \frac{\widehat{\Sigma}_\ell(x) \subseteq \vartheta \subseteq \Theta(\ell)}{(\widehat{\Sigma}, \Theta) \models_\ell x : \vartheta} \\
\text{(E-fun)} \quad \frac{\forall (Z_1, R_1), \dots, (Z_r, R_r) : \bigwedge_{i=1}^r (Z_i, R_i) \in \vartheta_i \Rightarrow (F^\ell, \{F^\ell \rightarrow f^\ell(Z_1, \dots, Z_r)\}) \cup \bigcup_{i=1}^r R_i \in \vartheta \subseteq \Theta(\ell)}{(\widehat{\Sigma}, \Theta) \models_\ell f(E_1, \dots, E_r) : \vartheta}
\end{array}$$

Table 2: Analysis of terms  $(\widehat{\Sigma}, \Theta) \models_\ell E : \vartheta$ .

- *abstract data collection*  $\Theta : \mathcal{L} \rightarrow 2^{\widehat{\mathcal{V}}}$  that, for each node labelled  $\ell$ , approximates the set of abstract values that the node handles.

The syntax directed rules of Tables 2 and 3 specify when an analysis estimate is valid and they are almost in AFPL format [39]. For each term  $E$ , the judgement  $(\widehat{\Sigma}, \Theta) \models_\ell E : \vartheta$  expresses that  $\vartheta \in 2^{\widehat{\mathcal{V}}}$  approximates the set of values that  $E$  may evaluate to, given the component  $\widehat{\Sigma}_\ell$  of the abstract store  $\widehat{\Sigma}$ . A sensor identifier and a value evaluate to the set  $\vartheta$ , provided that their abstract representations belong to  $\vartheta$  (rules (E-sen) and (E-val)). This abstract representation is a grammar made of a non-terminal symbol whose production generates a tree with a single node. For example, the abstract value for a sensor  $i$  is  $(I^\ell, \{I^\ell \rightarrow i^\ell\})$  that represents a grammar with the initial symbol is  $I^\ell$  that only generates the tree  $i^\ell$ . Similarly in rule (E-var) where a variable  $x$  evaluates to  $\vartheta$ , if this includes the set of values bound to  $x$  in  $\widehat{\Sigma}_\ell$ . The rule (E-fun) analyses the application of a  $r$ -ary function  $f$  to produce the set  $\vartheta$ . To do that (i) for each term  $E_i$ , it finds the sets  $\vartheta_i$ , and (ii) for all  $r$ -tuples of values  $(\hat{v}_1, \dots, \hat{v}_r)$  in  $\vartheta_1 \times \dots \times \vartheta_r$ , it checks if  $\vartheta$  includes the grammars with distinct symbol  $F^\ell$  generating the trees rooted in  $f^\ell$  with subtrees  $\hat{v}_1, \dots, \hat{v}_r$ . Also, in all the rules for terms, we require that the abstract data collection  $\Theta(\ell)$  includes all the abstract values in  $\vartheta$ .

In the analysis of nodes we focus on which values can flow through the network and which can be assigned to variables. The judgements have the form  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and are defined by the rules in Table 3. The rules for the inactive node (N-nil) and for parallel composition (N-par) are standard, as well as the rules (B-nil) and (B-par) for node components. The rule (N-node) for a single node  $\ell : [B]$  requires that its component  $B$  is analysed, with the further judgment  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B$ , where  $\ell$  is the label of the enclosing node. The rule (B-store) connects actual stores  $\Sigma$  with abstract ones  $\widehat{\Sigma}$  and it requires the locations of sensors to contain the corresponding abstract values. The rule (B-sen) for sensors does not inspect their form, because we are only interested in who will use their values and this information can be retrieved by the abstract store. The same happens in rule (B-act) for actuators that in our model are passive entities. Indeed, they obey commands of control processes and to track the activities of actuators it thus suffices to consider the issued commands.

The rules for processes are in the lower part of Table 3, and all require that an estimate is valid for the immediate sub-processes. The rule (P-out) for  $r$ -ary multi-output (i) finds the sets  $\vartheta_i$ , for each term  $E_i$ ; and (ii) for all  $r$ -tuples of values  $(\hat{v}_1, \dots, \hat{v}_r)$  in  $\vartheta_1 \times \dots \times \vartheta_r$ , it

$\frac{\text{(N-nil)}}{\widehat{\Sigma}, \kappa, \Theta \models 0}$	$\frac{\text{(N-par)}}{\widehat{\Sigma}, \kappa, \Theta \models N_1 \wedge \widehat{\Sigma}, \kappa, \Theta \models N_2}{\widehat{\Sigma}, \kappa, \Theta \models N_1 \mid N_2}$	$\frac{\text{(N-node)}}{\widehat{\Sigma}, \kappa, \Theta \models_\ell B}{\widehat{\Sigma}, \kappa, \Theta \models \ell : [B]}$	$\frac{\text{(B-nil)}}{\widehat{\Sigma}, \kappa, \Theta \models_\ell 0}$
$\frac{\text{(B-store)}}{\forall i \in \mathcal{I}_\ell. (I^\ell, \{I^\ell \rightarrow i^\ell\}) \in \widehat{\Sigma}_\ell(i)}{\widehat{\Sigma}, \kappa, \Theta \models_\ell \Sigma}$	$\frac{\text{(B-sen)}}{\widehat{\Sigma}, \kappa, \Theta \models_\ell S}$	$\frac{\text{(B-act)}}{\widehat{\Sigma}, \kappa, \Theta \models_\ell A}$	$\frac{\text{(B-par)}}{\widehat{\Sigma}, \kappa, \Theta \models_\ell B_1 \wedge \widehat{\Sigma}, \kappa, \Theta \models_\ell B_2}{\widehat{\Sigma}, \kappa, \Theta \models_\ell B_1 \parallel B_2}$
$\frac{\text{(P-out)}}{\bigwedge_{i=1}^k (\widehat{\Sigma}, \Theta) \models_\ell E_i : \vartheta_i \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_\ell P \wedge \forall \hat{v}_1, \dots, \hat{v}_r : \bigwedge_{i=1}^r \hat{v}_i \in \vartheta_i \Rightarrow \forall \ell' \in L : (\ell, \langle \langle \hat{v}_1, \dots, \hat{v}_r \rangle \rangle) \in \kappa(\ell')}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell \langle \langle E_1, \dots, E_r \rangle \rangle \triangleright L.P}$			
$\frac{\text{(P-in)}}{\bigwedge_{i=1}^j (\widehat{\Sigma}, \Theta) \models_\ell E_i : \vartheta_i \wedge \forall (\ell_2, \langle \langle \hat{v}_1, \dots, \hat{v}_r \rangle \rangle) \in \kappa(\ell_1) : \text{Comp}(\ell_2, \ell_1) \Rightarrow \left( \bigwedge_{i=j+1}^k \hat{v}_i \in \widehat{\Sigma}_{\ell_1}(x_i) \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_1} P \right)}{(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_1} (E_1, \dots, E_j; x_{j+1}, \dots, x_r).P}$			
$\frac{\text{(P-ass)}}{\forall \hat{v} \in \vartheta \Rightarrow \hat{v} \in \widehat{\Sigma}_\ell(x) \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_\ell P}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell x := E.P}$	$\frac{\text{(P-cond)}}{(\widehat{\Sigma}, \Theta) \models_\ell E : \vartheta \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_\ell P_1 \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_\ell P_2}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell E?P_1 : P_2}$		
$\frac{\text{(P-act)}}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell P}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell \langle j, \gamma \rangle.P}$	$\frac{\text{(P-rec)}}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell P}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell \mu h. P}$	$\frac{\text{(P-var)}}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell P}{(\widehat{\Sigma}, \kappa, \Theta) \models_\ell h}$	

Table 3: Analysis of nodes  $(\widehat{\Sigma}, \kappa, \Theta) \models N$ , and of node components  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B$ .

checks if they belong to  $\kappa(\ell' \in L)$ , i.e. if these  $r$ -tuples of values can be received by the nodes with labels in  $L$ . In the rule (P-in) for input the terms  $E_1, \dots, E_j$  are used for matching values sent on the network: this rule (i) checks whether the first  $j$  terms have valid estimates  $\vartheta_i$ ; and (ii) for each message  $(\ell_2, \langle \langle \hat{v}_1, \dots, \hat{v}_j, \hat{v}_{j+1}, \dots, \hat{v}_r \rangle \rangle)$  in  $\kappa(\ell_1)$  (i.e. in any message predicted to be receivable by the node with label  $\ell_1$ ), it checks that the values  $\hat{v}_{j+1}, \dots, \hat{v}_r$  are included in the estimates for the variables  $x_{j+1}, \dots, x_r$ , provided that the two nodes can communicate ( $\text{Comp}(\ell_2, \ell_1)$ ). The rule (P-ass) for assignment requires that  $\widehat{\Sigma}_\ell(x)$  includes all the values  $\hat{v}$  in  $\vartheta$ , the estimate for  $E$ . The rule (P-cond) is as expected and the rule (P-act) is trivial. Finally, the rules (P-rec) and (P-var) for iteration are standard, where to save notation, we assumed that each variable  $h$  is uniquely bound to the body  $P$ .

To show how our analysis works, consider again the example in Section 2 and the process  $P_{cp} = \mu h.(z := 1).(z' := \text{noiseRed}(z)).\langle z' \rangle \triangleright \{\ell_a\}$ .  $h$ . Moreover, let  $\iota$  and  $\nu$  be

$$\iota = (\mathcal{I}^{cp}, \{\mathcal{I}^{cp} \rightarrow 1^{cp}\}) \quad (4.1)$$

$$\nu = (\text{NOISERED}^{cp}, \{\text{NOISERED}^{cp} \rightarrow \text{noiseRed}^{cp}(\mathcal{I}^{cp}), \mathcal{I}^{cp} \rightarrow 1^{cp}\}) \quad (4.2)$$



abstracting values of the camera  $S_{cp}$  and its elaboration, respectively. Every valid CFA estimate must include at least the following entries:

$$\begin{aligned} (a) \widehat{\Sigma}_{\ell_{cp}}(z) \supseteq \{\iota\} & & (b) \widehat{\Sigma}_{\ell_{cp}}(z') \supseteq \{\iota, \nu\} \\ (c) \Theta(\ell_{cp}) \supseteq \{\iota, \nu\} & & (d) \kappa(\ell_a) \supseteq \{(\ell_{cp}, \langle\langle \nu \rangle\rangle)\} \end{aligned}$$

Indeed, all the following checks must succeed:

- $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_{cp}} \mu h.(z := \perp).(z' := \text{noiseRed}(z)).\langle\langle z' \rangle\rangle \triangleright \{\ell_a\}.h$  because
- $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_{cp}} (z := \perp).(z' := \text{noiseRed}(z)).\langle\langle z' \rangle\rangle \triangleright \{\ell_a\}$ , that in turn holds
- because (i)  $\iota$  is in  $\widehat{\Sigma}_{\ell_{cp}}(z)$  by (a) ( $(\widehat{\Sigma}, \Theta) \models_{\ell} \perp : \vartheta \ni \iota$ ); and because
  - (ii)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_{cp}} (z' := \text{noiseRed}(z)).\langle\langle z' \rangle\rangle \triangleright \{\ell_a\}$ , that in turn holds
- because (i)  $\nu$  is in  $\widehat{\Sigma}_{\ell_{cp}}(z')$  by (b) since
  - $(\widehat{\Sigma}, \Theta) \models_{\ell_{cp}} \text{noiseRed}(z) : \vartheta \ni \nu$ ; and because
  - (ii)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_{cp}} \langle\langle z' \rangle\rangle \triangleright \{\ell_a\}$  that holds because  $(\ell_{cp}, \langle\langle \nu \rangle\rangle)$  is in  $\kappa(\ell_a)$  by (d).

The precision of the CFA above can be refined by replacing the abstract store  $\widehat{\Sigma}$  with the pair  $\widehat{\Sigma}_{in}, \widehat{\Sigma}_{out}$ . This extension requires a more verbose specification of the rules for accurately handling the store updates, similarly to the treatment of side effects in [40]. We can obtain a further improvement of the precision by making the analysis more context-sensitive. In particular, an additional component can record the sequence of choices made in conditionals while traversing the node under analysis. One can thus obtain better approximations of the store or detect causal dependencies among the data sent by sensors and the actions carried out by actuators, as well as casuality among nodes.

**4.3. Correctness of the analysis.** Our CFA respects the operational semantics. The proof of this fact benefits from an instrumented denotational semantics for expressions. For that we will introduce a set of trees  $\widehat{\mathcal{T}}$ , ranged over by  $\hat{t}, \hat{t}', \dots$ , built over the ranked alphabet  $\mathbb{T}$  introduced above. The values of expressions become now pairs  $\langle v, \hat{t} \rangle$ , and the store and its updates are accordingly extended. The instrumented local store then becomes  $\Sigma_{\ell}^i : \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell} \rightarrow \mathcal{V} \times \widehat{\mathcal{T}}$ . We also endow  $\Sigma_{\ell}^i$  with an undefined value  $\perp$  for recording when a sensor or a variable is not initialised. Accordingly, we assume that the semantic function  $\text{measure}(i, \mathcal{E})$  returns a pair  $(v, i^{\ell})$ . Finally, when no ambiguity arises, we shall overload the symbol  $v$  to also denote the instrumented values.

The formal definition of the instrumented denotational semantics follows, where we indicate with  $\downarrow_i$  the projection on the  $i^{\text{th}}$  component of the pair.

$$\begin{aligned} \llbracket v \rrbracket_{\Sigma_{\ell}^i}^i &= (v, v^{\ell}) \\ \llbracket i \rrbracket_{\Sigma_{\ell}^i}^i &= \Sigma_{\ell}^i(i) \\ \llbracket x \rrbracket_{\Sigma_{\ell}^i}^i &= \Sigma_{\ell}^i(x) \\ \llbracket f(E_1, \dots, E_r) \rrbracket_{\Sigma_{\ell}^i}^i &= (f(\llbracket E_1 \rrbracket_{\Sigma_{\ell}^i \downarrow_1}^i, \dots, \llbracket E_r \rrbracket_{\Sigma_{\ell}^i \downarrow_1}^i), f^{\ell}(\llbracket E_1 \rrbracket_{\Sigma_{\ell}^i \downarrow_2}^i, \dots, \llbracket E_r \rrbracket_{\Sigma_{\ell}^i \downarrow_2}^i)) \end{aligned}$$

Since the CFA only considers the second component of the extended store, we need to define when the concrete and the abstract stores agree.

**Definition 1.** Given a concrete store  $\Sigma_{\ell}^i$  and an abstract store  $\widehat{\Sigma}_{\ell}$ , we say that they *agree*, in symbols  $\Sigma_{\ell}^i \bowtie \widehat{\Sigma}_{\ell}$ , if and only if for all  $w \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  either  $\Sigma_{\ell}^i(w) = \perp$  or there exists  $\widehat{G} \in \widehat{\Sigma}_{\ell}(w)$  such that  $(\Sigma_{\ell}^i(w))_{\downarrow_2} \in \text{Lang}(\widehat{G})$ .

Just to give an intuition, suppose that the expression  $E$  is such that  $\llbracket E \rrbracket_{\Sigma_\ell}^i = (v, v^\ell)$ . Then, the assignment  $x := E$  will result in the updated store  $\Sigma_\ell^i\{(v, v^\ell)/x\}$ . Clearly, the standard semantics of expressions used in Table 1 is obtained by the projection on the first component of the instrumented one. In our running example, the assignment  $z' := \text{noiseRed}(z)$  of the process  $P_{cp}$  stores the pair  $(v, \text{noiseRed}^{\ell_{cp}}(1^{\ell_{cp}}))$  made of the actual value  $v$  and of its abstract counterpart.

The following theorem establishes the correctness of our CFA in that its valid estimates are preserved under reduction steps.

**Theorem 4.1** (Subject reduction). *Given a system of nodes  $N$ , if  $(\widehat{\Sigma}, \kappa, \Theta) \models N; N \xrightarrow[\varepsilon']{\varepsilon} N'$  and  $\forall \Sigma_\ell^i$  in  $N$  it is  $\Sigma_\ell^i \bowtie \widehat{\Sigma}_\ell$ , then  $(\widehat{\Sigma}, \kappa, \Theta) \models N'$  and  $\forall \Sigma_\ell^{i'}$  in  $N'$  it is  $\Sigma_\ell^{i'} \bowtie \widehat{\Sigma}_\ell$ .*

We also prove that the set of valid estimates to the specification in Table 3 is never empty and that a minimal one always exists. This is because estimates form a Moore family  $\mathcal{M}$ , i.e. a set with a greatest element ( $\sqcup \emptyset$ ) and a least element ( $\sqcap \mathcal{M}$ ).

**Theorem 4.2** (Existence of estimates). *Given  $N$ , its valid estimates form a Moore family.*

We now illustrate the role that the components of a valid estimate play in predicting the behaviour of the analysed system. The following corollary follows from the fact that the analysis respects the operational semantics, as stated in Theorem 4.1. The first item below makes it evident that our analysis determines whether the value of a term may indeed be used along the computations of a system, and clarifies the role of the component  $\Theta$ ; the second item guarantees that  $\kappa$  predicts all the possible inter-node communications. In the statement of this corollary we use the following notation for reductions, where we omit writing the environments for readability. Let  $N \xrightarrow{E_1, \dots, E_r}_\ell N'$  denote a reduction in which all the expressions  $E_i$  are evaluated at node  $\ell$  and let  $N \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N'$  be a reduction in which the message sent by node  $\ell_1$  is received by node  $\ell_2$ .

**Corollary 4.3.** *Assume that  $(\widehat{\Sigma}, \kappa, \Theta) \models N$ , then*

- (1) *if  $N \xrightarrow{E_1, \dots, E_r}_\ell N'$  then  $\forall k \in [0, n]$  there exists  $\widehat{G} \in \Theta(\ell)$  such that  $(\llbracket E \rrbracket_{\Sigma_\ell}^i)_{\downarrow_2} \in \text{Lang}(\widehat{G})$ ;*
- (2) *if  $N \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N'$  then it holds  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$ , and  $\forall i \in [0, r]$  there exists  $\widehat{G} \in \hat{v}_i$  such that  $v_{i\downarrow_2} \in \text{Lang}(\widehat{G})$ .*

Back again to our example at the end of sub-section 4.2, consider the assignment  $z := 1$  in the process  $P_{cp}$ . We have  $(\llbracket 1 \rrbracket_{\Sigma_{\ell_{cp}}^1}^1)_{\downarrow_2} = 1^{\ell_{cp}}$  where  $v$  is the actual value received by the first sensor, and that our analysis computes  $\iota \in \theta(\ell_{cp})$ , where  $\iota$  is defined in (4.1). It is immediate to see that  $1^{\ell_{cp}} \in \text{Lang}(\iota)$ . A similar reasoning can be done also for messages from  $\ell_{cp}$  to  $\ell_a$ . Indeed, we have  $(\llbracket z' \rrbracket_{\Sigma_{\ell_{cp}}^1}^1)_{\downarrow_2} = \text{noiseRed}^{\ell_{cp}}(1^{\ell_{cp}})$  and that our analysis computes  $(\ell_{cp}, \langle\langle \nu \rangle\rangle) \in \kappa(\ell_a)$ , where  $\nu$  is defined in (4.2). It is immediate to see that  $\text{noiseRed}^{\ell_{cp}}(1^{\ell_{cp}}) \in \text{Lang}(\nu)$ .

**4.4. Some applications of the analysis.** To give an idea on how the outcome of the analysis can be used to detect where and how data are manipulated and how messages flow in a system, we introduce below a couple of simple properties.

A designer could be interested in checking whether a value taken by a specific sensor of a node will be used as an ingredient of the data of a different node. This is formalised as follows.

**Definition 2.** Let  $N$  be the node  $\ell : [B]$  containing the sensor  $i$ .

The sensor  $i$  is an *ingredient* of a node  $N'$  with label  $\ell'$  if and only if  $N' \xrightarrow{E_1, \dots, E_r}_{\ell'} N''$  and there exists  $k \in [1, r]$  such that  $(\llbracket E \rrbracket_{\Sigma_i}^i)_{\downarrow 2}$  is a tree with a leaf  $i^\ell$ .

The following property is an immediate consequence of Corollary 4.3.

**Proposition 4.4.** Let  $i$  be a sensor of the node  $N$  with label  $\ell$ ; let  $N'$  be a node with label  $\ell'$ ; and let  $(\widehat{\Sigma}, \kappa, \Theta) \models N \mid N'$ .

The sensor  $i$  is an ingredient of  $N'$  if there exists  $\widehat{G} \in \Theta(\ell')$  and a tree in its language with a leaf  $i^\ell$ .

Consider again our example at the end of sub-section 4.2, for which estimates are such that  $\Theta(\ell_a) \supseteq \{\nu\}$  (because  $\kappa(\ell_a) \supseteq \{(\ell_{cp}, \langle\langle \nu \rangle\rangle)\}$ ), showing that the sensor  $1^{cp}$  is an ingredient of the node  $N_a$ .

Note that the schema above can be used to verify other properties by inspecting the component  $\Theta$ . It suffices to express properties of tree languages and check them with the standard tools, e.g. tree automata model checking.

Our next example works over the problem of checking the robustness of the communications in the smart street light example discussed at the end of Section 2, where we assumed the lamp post  $\ell_q$  to be our of order.

Consider the following abstract values, each for every  $p$ :

$$\nu_p^c = (\text{CAR}^p, \{\text{CAR}^p \rightarrow \text{car}^p\}) \quad \nu_p^w = (\text{TRUE}^p, \{\text{TRUE}^p \rightarrow \text{true}^p\})$$

the first is used to signal the presence of a car, the second one represents a boolean for the pedestrian. Every valid CFA estimate for our example includes in the component  $\kappa$  the following entries for all lamp posts  $p$ :

$$\kappa(\ell_p) \supseteq \{(\ell_{p-1}, \langle\langle \nu_{p-1}^c \rangle\rangle), (\ell_{p-1}, \langle\langle \nu_{p-1}^w \rangle\rangle), (\ell_{p+1}, \langle\langle \nu_{p+1}^w \rangle\rangle)\}$$

In particular, this holds also for the lamp post  $\ell_{q+1}$  and the designer can check that the messages  $(\ell_q, \langle\langle \nu_q^c \rangle\rangle)$ ,  $(\ell_q, \langle\langle \nu_q^w \rangle\rangle)$  are never received because the lamp post  $\ell_q$  is out of order.

If needed, the designer can make this situation more explicit, by encoding the fault of  $\ell_q$  in the compatibility function. To this aim, it suffices to modify the function only for the faulty lamp post, by putting  $\text{Comp}(\ell_q, \ell_p) = \text{false}$  for all lamp posts. Then, the designer can re-apply the analysis and obtain that  $(\ell_q, \langle\langle \nu_q^c \rangle\rangle)$ ,  $(\ell_q, \langle\langle \nu_q^w \rangle\rangle) \notin \kappa(\ell_{q+1})$ , from which deducing that these messages will never be received by the lamp posts from position  $q + 1$  onwards.

## 5. CFA AT WORK: VERIFYING SECURITY POLICIES

Security is a crucial issue in IoT since it “deals not only with huge amount of sensitive data (personal data, business data, etc.) but also has the power of influencing the physical environment with its control abilities” [30]. Traditional security risks are magnified [5] in the IoT world, where opportunities for potential breaches dramatically increase hand in hand with the number of connected *things*. Actually, security concerns both the physical and the logical facets of cyber-physical systems and poses manifold challenges. However, these aspects have not received much attention so far. Counter-measures against a physical

attack are very hard to implement, sensors are often in not protected locations and can thus easily be damaged and their transmissions disrupted. Yet more standard, security of the logical components raises new issues, mainly because IoT systems must perform all in a lightweight way.

Here we extend our proposal with encryption and decryption constructs for representing and handling some aspects of security from the side of applications. Encryption is indeed the traditional mechanism to provide security, but it is very expensive in a setting where energy, computation and storage are rather limited. Our CFA may help IoT designers in singling out the pieces of information to be kept confidential and in detecting whether the security assets are protected inside the system. In addition, we will show that a static check suffices to detect if certain data are propagated within a system of nodes according to a given policy. In the first part of this section we propose a general schema that captures a family of security and access control policies. We then instantiate the schema to a couple of classical security notions and a more sophisticated policy that controls how information is propagated.

In particular, we show how our CFA can address data security, by detecting leakages, once values and data are partitioned in confidential and public. The designer of a system of nodes can then suitably protect sensitive information, by only encrypting their relevant parts, typically the values coming from sensors or from data aggregations. In addition, we re-cast the classical no-read-up/no-write down in our framework. Indeed, we can then statically check if information is exchanged between nodes in the right direction with respect to their level of clearance. Finally, we instantiate the general schema with a finer policy that restricts the propagation of certain confidential data only within a specific sub-system. We show that our CFA suffices to detect whether some confidential information unintentionally reaches a node that should not get it. Remarkably, verifying the above properties, and others, is done on the estimates of our CFA that are computed once and for all.

We are here assuming a symmetric encryption schema, which is the one commonly used in IoT systems because less energy consuming; however, using public key encryption will only require slight changes in the definitions below. Formally, the encryption of a message  $m$  under a key  $k$  can be defined as the result of an aggregation function  $encrypt(m, k)$  and the decryption as the application of its inverse  $encrypt^{-1}(x, k)$ . The machinery developed so far would then formally work as it is. However the importance of security calls for an explicit treatment of its aspects, described in this section.

Our current analysis is restricted to the system of nodes under investigation, and does not consider the possible presence of active malicious attackers that can intervene in communications by injecting forged messages. We do not feel these restrictions too heavy. One reason is that in many real IoT systems, e.g. those developed with Zigbee [46], keys are seldom exchanged in communications. In this real world systems cryptographic keys, if present at all, are often fixed once and for all and exchanged in a secure manner when the system starts its setup or upon a reboot. Also, it is not difficult to handle the case when the design of the system benefits from a key distribution. Indeed, one can re-use existent techniques for doing that, e.g. based on CFA [8, 45], through which also active attackers are implicitly included.

**5.1. Extending the language and the analysis.** We extend here the syntax of processes and terms with encryption and decryption primitives as in L<sub>Y</sub>SA, taking for simplicity a

symmetric encryption schema. As argued above, we feel free to assume as given a finite set  $\mathcal{K}$  of secret keys owned by nodes, previously exchanged in a secure way.

Terms now include also encryptions written

$$\{E_1, \dots, E_r\}_k$$

the result of which comes from encrypting the values of the expressions  $E_i$ , for  $i \in [1, r]$ , under the shared key  $k$  in  $\mathcal{K}$ .

The syntax of processes now has also the following construct

$$\text{decrypt } E \text{ as } \{E_1, \dots, E_j; x_{j+1}, \dots, x_r\}_k \text{ in } P$$

This process receives a message encrypted with the shared key  $k \in \mathcal{K}$ . Also in this case we use pattern matching, as explained below.

The new decryption construct requires the following semantic rule:

(Decrypt)

$$\llbracket E \rrbracket_\Sigma = \{v_1, \dots, v_r\}_k \wedge \bigwedge_{i=1}^j v_i = \llbracket E'_i \rrbracket_\Sigma$$

$$\Sigma \parallel \text{decrypt } E \text{ as } \{E'_1, \dots, E'_j; x_{j+1}, \dots, x_r\}_k \text{ in } P \parallel B \xrightarrow[\mathcal{E}]{\mathcal{E}} \Sigma\{v_{j+1}/x_{j+1}, \dots, v_r/x_r\} \parallel P \parallel B$$

In this inference rule the encryption  $\{v_1, \dots, v_r\}_k$ , resulting from  $E$ , has to match against the pattern in  $\text{decrypt } E \text{ as } \{E'_1, \dots, E'_j; x_{j+1}, \dots, x_r\}_k \text{ in } P$ , i.e. the first  $j$  values  $v_i$  must match those of the corresponding  $E'_i$ . In addition the keys must be the same (this models *perfect* symmetric cryptography). When all the above holds, the values of the remaining expressions are bound to the corresponding variables.

In order to extend the analysis for each node  $\ell$ , we first add to the alphabet  $\mathbb{T}$  the following new symbols, with elements:

- $enc_r^\ell$ , each with rank  $r$  ranging from 2 to  $m$ , where  $m - 1$  is the maximum arity of the encrypted terms used in the system;
- $k_i$ , for each key  $k_i \in \mathcal{K}$ .

As done above, we assume that there is a non-terminal for each of the new terminals, written with capital letters. It is convenient to define the following auxiliary function  $D$  for extracting the ordered list of the grammars concerning the abstract sub-terms of an encryption under the key  $k$  (otherwise it returns the empty list):

$$D((\text{ENC}_r^\ell, R), k) = [(Z_i, R_i) \mid \exists \text{ENC}_r^\ell \rightarrow enc_r^\ell(Z_1, \dots, Z_r, K), K \rightarrow k \in R \wedge R_i = D^*(Z, R)]$$

where  $D^*(Z, R)$  extracts the productions in  $R$  that are used to derive trees from  $Z$  and is defined below

$$D^*(Z, R) = \{Z \rightarrow \hat{t} \in R \mid \hat{t} \in \mathbb{T}\} \cup \bigcup_{Z \rightarrow M \in R} \{D^*(W, R) \mid W \text{ occurs in } M \in \mathbb{N} \cup \mathbb{T}\}$$

To better understand how the function  $D$  works, consider the node  $N'_a$  of the example in Section 2. The analysis tells us that the variable  $y$  could take the abstract value  $\epsilon$  defined as follows

$$(\text{ENC}_1^{cp}, \{ \text{ENC}_1^{cp} \rightarrow enc_1^{cp}(\text{NOISERED}^{cp}, K), K \rightarrow k, \text{NOISERED}^{cp} \rightarrow \text{NoiseRed}^{cp}(\mathcal{I}^{cp}), \mathcal{I}^{cp} \rightarrow 1^{cp} \}).$$

The function  $D$  applied to  $\epsilon$  extracts the grammar approximating the enhanced version of the picture taken by the sensor 1 in the node  $N_{cp}$ :

$$(\text{NOISERED}^{cp}, \{\text{NOISERED}^{cp} \rightarrow \text{noiseRed}^{cp}(\mathcal{I}^{cp}), \mathcal{I}^{cp} \rightarrow 1^{cp}\}).$$

We now introduce the following new rules for analysing encrypted terms and decryption processes.

$$\begin{array}{c}
\text{(E-enc)} \\
\frac{\bigwedge_{i=1}^r (\widehat{\Sigma}, \Theta) \models_{\ell} E_i : \vartheta_i \wedge \forall (Z_1, R_1), \dots, (Z_r, R_r) : \bigwedge_{i=1}^r (Z_i, R_i) \in \vartheta_i \Rightarrow \\
\text{(ENC}_r^{\ell}, \{\text{ENC}_r^{\ell} \rightarrow \text{enc}_r^{\ell}(Z_1, \dots, Z_r, K), K \rightarrow k\} \cup \bigcup_{i=1}^r R_i) \in \vartheta \subseteq \Theta(\ell)}{(\widehat{\Sigma}, \Theta) \models_{\ell} \{E_1, \dots, E_r\}_k : \vartheta} \\
\text{(P-dec)} \\
\frac{(\widehat{\Sigma}, \Theta) \models_{\ell} E : \vartheta \wedge \bigwedge_{i=1}^j (\widehat{\Sigma}, \Theta) \models_{\ell} E_i : \vartheta_i \wedge \\
\forall \hat{v} \in \vartheta \text{ s.t. } D(\hat{v}, k) = [\hat{v}_1, \dots, \hat{v}_r] \Rightarrow \left( \bigwedge_{i=j+1}^r \hat{v}_i \in \widehat{\Sigma}_{\ell}(x_i) \wedge (\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} P \right)}{(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \text{decrypt } E \text{ as } \{E_1, \dots, E_j; x_{j+1}, \dots, x_r\}_k \text{ in } P}
\end{array}$$

Similarly to the rule for evaluating function applications, (E-enc) analyses an encryption term and produces the set  $\vartheta$ . To do that (i) for each term  $E_i$ , it finds the sets  $\vartheta_i$ , and (ii) for all  $r$ -tuples of values  $(\hat{v}_1, \dots, \hat{v}_r)$  in  $\vartheta_1 \times \dots \times \vartheta_r$ , it checks if  $\vartheta$  includes the grammar with the production for the single node  $k$ , the key, and distinct symbol  $\text{ENC}_r^{\ell}$ , with arity  $r$ , generating the trees rooted in  $\text{enc}_r^{\ell}$  with subtrees  $\hat{v}_1, \dots, \hat{v}_r$ . The premises of the rule (P-dec) (i) check if each  $\vartheta_i$  approximates the values of each term  $E_i$  involved in the matching; (ii) inspect the approximation  $\vartheta$  of  $E$  to be decrypted and from each  $\hat{v}$  in this abstract value it extracts the set of ordered lists  $L = [\hat{v}_1, \dots, \hat{v}_r]$  of the grammars associated with the sub-terms of  $E$  through the auxiliary function  $D$ ; (iii) check that the presence of such a list  $L$  implies that the values  $\hat{v}_i$  that at run time are bound to  $x_i$  ( $i \in [j+1, r]$ ) are correctly predicted (i.e. belong to  $\widehat{\Sigma}_{\ell}(x_i)$ ), and in addition that the estimate in hand is valid for the continuation  $P$  (note that this last check is only done if there exist a non-empty list  $L$ ). If all the above holds, the decryption process is correctly analysed.

In order to establish the subject reduction for proving the correctness of this extension to our CFA, we extend the instrumented denotational semantics for terms in the following expected way:

$$[[\{E_1, \dots, E_r\}_k]_{\Sigma_i}^i] = (\{[[E_1]]_{\Sigma_{i \downarrow 1}}^i, \dots, [[E_r]]_{\Sigma_{i \downarrow 1}}^i\}_k, \text{enc}_r^{\ell}([E_1]_{\Sigma_{i \downarrow 2}}^i, \dots, [E_r]_{\Sigma_{i \downarrow 2}}^i, k))$$

Now we simply assert that the theorems and the corollary of the previous sections still hold, without formally writing them, as nothing changes in their statements. Their proofs have already been included in those of the original versions.

**5.2. Controlling data propagation.** We now introduce our general schema to express security policies and a way of checking them at static time. As said above, we provide the designer with ways to classify data and nodes and to regulate how data are propagated within the system. In the next sub-sections we will illustrate this schema by instantiating it on three specific cases.

In order to classify data, the designer associates with them a security tag, taken from a finite set  $\mathfrak{D}$ . Then, it defines a pair of functions that accordingly assign tags to abstract values. As expected, all the trees generated by a grammar must have the same tag that is also assigned to the grammar.

**Definition 3** (Tagging data). Given a finite set  $\mathfrak{D}$  of *tags*, the functions

$$\mathfrak{T}_{\mathfrak{D}} : \widehat{\mathcal{T}} \rightarrow \mathfrak{D} \qquad \mathfrak{T}_{\mathfrak{G}} : \widehat{\mathcal{V}} \rightarrow \mathfrak{D}$$

are *tagging functions* if and only if, given  $\widehat{G} \in \widehat{\mathcal{V}}$ , for all  $\hat{t} \in \text{Lang}(\widehat{G})$  the following condition holds

$$\mathfrak{T}_{\mathfrak{D}}(\hat{t}) = \mathfrak{T}_{\mathfrak{G}}(\widehat{G})$$

The other element of the schema is a policy  $\mathfrak{P}$ . In our case, it is simply defined as a predicate over tags and labels. Intuitively, a designer dictates through a policy if data with a specific tag can be exchanged between two nodes and how, e.g. encrypted or as clear text. Our schema is formalised below.

**Definition 4** (Data propagation policies). Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ ; let  $\mathfrak{D}$  be a finite set of tags; let  $\mathfrak{T}_{\mathfrak{D}} : \widehat{\mathcal{T}} \rightarrow \mathfrak{D}$ ; and let  $\mathfrak{P} \subseteq \mathfrak{D} \times \mathcal{L} \times \mathcal{L}$  be a data propagation policy.

Then  $N$  *enjoys*  $\mathfrak{P}$  if and only if  $N \rightarrow^* N'$  and for all  $\ell_1, \ell_2 \in \mathcal{L}$  there is no transition such that  $N' \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N''$  and  $\mathfrak{P}(\mathfrak{T}_{\mathfrak{D}}(v_{i \downarrow 2}), \ell_1, \ell_2)$  does not hold, for some  $i \in [1, r]$ .

The following theorem guarantees that the component  $\kappa$  of the analysis can be inspected to statically check if a given system of nodes enjoys the wanted security policy.

**Theorem 5.1** (Well propagation). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ ; let  $\mathfrak{D}$  be a finite set of tags; let  $\mathfrak{T}_{\mathfrak{D}}, \mathfrak{T}_{\mathfrak{G}}$  be a pair of tagging functions; and let  $\mathfrak{P} \subseteq \mathfrak{D} \times \mathcal{L} \times \mathcal{L}$ .*

*Then  $N$  enjoys  $\mathfrak{P}$  if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then  $\forall i \in [1, r]$  it holds  $\mathfrak{P}(\mathfrak{T}_{\mathfrak{G}}(\hat{v}_i), \ell_1, \ell_2)$ .

Note that a more general version of the above can be obtained by turning the set  $\mathfrak{D}$  into a finite complete lattice  $(\mathfrak{D}, \sqsubseteq)$ ; requiring that the tagging functions satisfy  $\mathfrak{T}_{\mathfrak{D}}(\hat{t}) \sqsubseteq \mathfrak{T}_{\mathfrak{G}}(\widehat{G})$  rather than equality; and imposing that  $\mathfrak{P}$  is monotone in its first argument.

The next two sub-sections instantiate the above schema on two classical security properties, confidentiality and no-read-up/no-write down, while the third sub-section instantiates the above schema the security property discussed in the example of Section 2. As we will see, confidentiality only requires classifying data and imposes no constraint on nodes. Instead, the second property neglects the tags of data and only considers clearance levels of nodes. The last one has constraints on both data and the way these are propagated within the system.

**Preventing Leakage.** As the first example of how the notion of well propagation is instantiated, we consider the common approach of identifying the sensitive content of information and of detecting possible disclosures. Hence, we partition values into security classes and prevent classified information from flowing in clear or to the wrong places. In the following, we assume that the designer defines the set  $\mathcal{S}_\ell$  containing the sensors of the node  $N$  with label  $\ell$  that have to be protected, i.e. the values of which are to be kept *secret*. This implicitly introduces also the set  $\mathcal{P}_\ell$  of *public* sensors as the complement of  $\mathcal{S}_\ell$ . In this case, the set of tags  $\mathfrak{D}$  is  $\{\text{secret}, \text{public}\}$ .

The abstract values are partitioned through the two tagging functions  $\mathcal{D}$  and  $\mathcal{S}$  defined below. The intuition behind them is that a single “drop” of *secret* turns to *secret* the (abstract) term. Of course there is the exception for encrypted data: what is encrypted is public, even if it contains secret components. Besides classifying sensors, one can obviously consider the case of aggregation functions whose result has to be kept secret.

**Definition 5.** Given the sets  $\mathcal{S}_\ell$  and  $\mathcal{P}_\ell$  of *secret* and of *public* sensors of the node  $N$  with label  $\ell$ , we define the pair of functions

- $\mathcal{D} : \hat{\mathcal{T}} \rightarrow \{\text{secret}, \text{public}\}$  as follows:

$$\mathcal{D}(i^\ell) = \begin{cases} \text{secret} & \text{if } i \in \mathcal{S}_\ell \\ \text{public} & \text{if } i \in \mathcal{P}_\ell \end{cases} \quad \mathcal{D}(v^\ell) = \text{public}$$

$$\mathcal{D}(\text{encr}_r^\ell(\hat{t}_1, \dots, \hat{t}_r, k)) = \text{public}$$

$$\mathcal{D}(f^\ell(\hat{t}_1, \dots, \hat{t}_r)) = \begin{cases} \text{secret} & \text{if } \exists \hat{t}_i \text{ s.t. } \mathcal{D}(\hat{t}_i) = \text{secret} \\ \text{public} & \text{otherwise} \end{cases}$$

- $\mathcal{S} : \hat{\mathcal{V}} \rightarrow \{\text{secret}, \text{public}\}$  as follows:

$$\mathcal{S}(\hat{v}) = \begin{cases} \text{secret} & \text{if } \exists (Z, R) \in \hat{v} \text{ s.t. } \mathcal{S}^*((Z, R)) = \text{secret} \\ \text{public} & \text{otherwise} \end{cases}$$

where  $\mathcal{S}^* : \mathbb{N} \times \mathbb{R} \rightarrow \{\text{secret}, \text{public}\}$  is:

$$\mathcal{S}^*((I^\ell, R)) = \begin{cases} \text{secret} & \text{if } i^\ell \in \mathcal{S}_\ell \\ \text{public} & \text{if } i^\ell \in \mathcal{P}_\ell \end{cases} \quad \mathcal{S}^*((V^\ell, R)) = \text{public}$$

$$\mathcal{S}^*((\text{ENC}_k^\ell, R)) = \text{public}$$

$$\mathcal{S}^*((F^\ell, \{F^\ell \rightarrow f^\ell(Z_1, \dots, Z_r)\} \cup R)) = \begin{cases} \text{secret} & \text{if } \exists (Z_i, R_i) \text{ s.t. } \mathcal{S}^*((Z_i, R_i)) = \text{secret} \\ \text{public} & \text{otherwise} \end{cases}$$

**Fact 5.2.**  $\mathcal{D}$  and  $\mathcal{S}$  are a pair of tagging functions.

Since our analysis computes information on the values exchanged during the communication, we can statically check whether a value, devised to be secret to a node  $N$ , is never sent to another node, similarly to [13]. We first give a dynamic characterisation of when a node  $N$  never discloses its secret values, i.e. when neither it nor any of its derivatives



can send a message that includes a secret value (recall that  $N \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N'$  stands for sending the message from  $\ell_1$  to  $\ell_2$ ).

**Definition 6.** Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ , and let  $\mathcal{S} = \{\mathcal{S}_\ell \mid \ell \in \mathcal{L}\}$  be the set of its secret sensors.

Then  $N$  has *no leaks* with respect to  $\mathcal{S}$  if and only if  $N \rightarrow^* N'$  and for all  $\ell_1, \ell_2 \in \mathcal{L}$  there is no transition  $N' \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N''$  such that  $\mathcal{D}(v_{i \downarrow \ell_2}) = \text{secret}$  for some  $i \in [1, r]$ .

The definition above instantiates the predicate  $\mathfrak{P}$  by imposing that the first argument has to be *public* and it simply ignores the labels of the nodes. The component  $\kappa$  allows us to statically detect when a system of nodes  $N$  has no leaks with respect to a given set of secret sensors, as an instance of Theorem 5.1.

**Theorem 5.3** (Confidentiality). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ , and let  $\mathcal{S} = \{\mathcal{S}_\ell \mid \ell \in \mathcal{L}\}$  be the set of its secret sensors. Then  $N$  has no leaks with respect to  $\mathcal{S}$  if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then  $\forall i \in [1, r]$   $\mathcal{S}(\hat{v}_i) = \text{public}$ .

Back to our running example of Section 2, we said that the pictures of cars are sensitive data, and that consequently one would like to keep them secret. To check this, we classify data coming from the sensor  $S_{cp}$  such that  $\mathcal{D}(1^{cp}) = \mathcal{S}(\iota) = \text{secret}$  (recall the definition of  $\iota$  in (4.1)). Accordingly to Definition 5, we also get  $\mathcal{D}(\text{noiseRed}^{cp}(1^{cp})) = \mathcal{S}(\nu) = \text{secret}$  (recall  $\nu$  from (4.2)). Hence, by simply inspecting  $\kappa$  we discover that the sensitive data of cars could be sent in clear to the street supervisor  $\ell_a$ , so violating secrecy: indeed  $\kappa(\ell_a) \supseteq \{(\ell_{cp}, \langle\langle \nu \rangle\rangle)\}$ . We illustrated an amended system in Section 2 that does not suffer from this problem.

**Communication policies.** We present a second example of instantiation of the general schema above. It considers the case when security is enforced by defining policies that rule information flows among nodes, by allowing some flows and forbidding others.

Below we consider the well-known *no read-up/no write-down* policy [4, 23]. It is based on a hierarchy of clearance levels for nodes, and it requires that a node classified at a high level cannot write any value to a node at a lower level, while the converse is allowed; symmetrically a node at low level cannot read data from one of a higher level.

For us, it suffices to classify the node labels with an assignment function  $level : \mathcal{L} \rightarrow \mathbf{L}$ , from the set of node labels to a given set of levels  $\mathbf{L}$ . We then introduce a condition for characterising the allowed and the forbidden flows: it suffices requiring that whenever a message is sent by to a node  $\ell_1$  to a node  $\ell_2$ , their levels satisfy the condition  $level(\ell_1) \leq level(\ell_2)$ , as defined below. We instantiate the schema as follows: we take a singleton set for  $\mathfrak{D}$ ; a pair of constant functions as tagging functions; and a predicate  $\mathfrak{P}$  that ignores the first argument and checks the required disequalities between levels.

**Definition 7.** Given an assignment function  $level$ , and a system of nodes  $N$  with labels in  $\mathcal{L}$ , then  $N$  *respects the levels* if and only if  $N \rightarrow^* N'$  and for all  $\ell_1, \ell_2 \in \mathcal{L}$  there is no transition such that  $N' \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N''$  and  $level(\ell_1) \leq level(\ell_2)$ .

Again the component  $\kappa$  of the analysis can be used to statically predict when the components of a system of nodes will communicate respecting the assigned levels of clearance, as an instance of Theorem 5.1.

**Theorem 5.4** (No-read-up/no-write-down). *Given an assignment function  $\text{level}$ , and a system of nodes  $N$  with labels in  $\mathcal{L}$ . Then  $N$  respects the levels if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle \hat{v}_1, \dots, \hat{v}_r \rangle) \in \kappa(\ell_2)$  then  $\text{level}(\ell_1) \leq \text{level}(\ell_2)$ .

More in general, we can constrain communication flows according to a specific policy, by stating whom a node is allowed to send (and/or from which it can receive) a message. It suffices to define a relation  $\mathbf{R} \subseteq \mathcal{L} \times 2^{\mathcal{L}}$ , in place of the relation used above that is based on the partial ordering between the levels of nodes.

**Selective data propagation.** We consider now a form of access control policy that confines exchange of a specific piece of information within a group of nodes in a system. An example of such a policy has been intuitively discussed at the end of Section 2. It dictates that the picture of an incoming car must only be sent to  $N_a$ , the supervisor of the camera, and to  $N_s$ , the supervisor of the lamp posts, but not to the lamp posts  $N_p$ , if the picture has not been anonymised.

In order to express the above policy in our general schema, we follow the pattern of the above examples. Assume to have the set  $\mathcal{R}_\ell$  containing the sensors of the node  $N$  with label  $\ell$ , the values of which are to be kept *confined*, together with its complement, the *open* values. In this case, the set  $\mathfrak{D}$  is  $\{\text{confined}, \text{open}\}$ .

The abstract values are partitioned through the two tagging functions  $\mathcal{C}$  and  $\mathcal{O}$  defined below. The intuition behind them is that a single “drop” of *confined* in a term makes such the whole, with the exception when the anonymisation function is applied: what is anonymised is *open*, even if it contains confined elements.

**Definition 8.** Let  $\mathcal{R}_\ell$  be the set of *confined* sensors of the node  $N$  with label  $\ell$ , and  $\tilde{\mathcal{F}} \subseteq \mathcal{F}$  be the set of anonymisation functions, possibly including encryptions. Then we define the following pair of functions

- $\mathcal{C} : \widehat{\mathcal{T}} \rightarrow \{\text{confined}, \text{open}\}$  as follows:

$$\mathcal{C}(i^\ell) = \begin{cases} \text{confined} & \text{if } i \in \mathcal{T}_\ell \\ \text{open} & \text{otherwise} \end{cases} \quad \mathcal{C}(v^\ell) = \text{open}$$

$$\mathcal{C}(f^\ell(\hat{t}_1, \dots, \hat{t}_r)) = \begin{cases} \text{confined} & \text{if } \exists \hat{t}_i \text{ s.t. } \mathcal{C}(\hat{t}_i) = \text{confined} \wedge f \notin \tilde{\mathcal{F}} \\ \text{open} & \text{otherwise} \end{cases}$$

- $\mathcal{O} : \widehat{\mathcal{V}} \rightarrow \{\text{confined}, \text{open}\}$  as follows:

$$\mathcal{O}(\hat{v}) = \begin{cases} \text{confined} & \text{if } \exists (Z, R) \in \hat{v} \text{ s.t. } \mathcal{O}^*((Z, R)) = \text{confined} \\ \text{open} & \text{otherwise} \end{cases}$$

where  $\mathcal{O}^* : \mathbb{N} \times \mathbb{R} \rightarrow \{\text{confined}, \text{open}\}$  is:

$$\mathcal{O}^*((I^\ell, R)) = \begin{cases} \text{confined} & \text{if } i^\ell \in \mathcal{T}_\ell \\ \text{open} & \text{otherwise} \end{cases} \quad \mathcal{O}^*((V^\ell, R)) = \text{open}$$

$$\mathcal{O}^*((F^\ell, \{F^\ell \rightarrow f^\ell(Z_1, \dots, Z_r)\} \cup R)) = \begin{cases} \text{confined} & \text{if } \exists (Z_i, R_i) \text{ s.t. } \mathcal{O}^*((Z_i, R_i)) = \text{confined} \wedge f \notin \tilde{\mathcal{F}} \\ \text{open} & \text{otherwise} \end{cases}$$

**Fact 5.5.**  $\mathcal{C}$  and  $\mathcal{O}$  are a pair of tagging functions.

The following definition formalises when a system of nodes propagates confined data only within the selected sub-system. For that, we let the set  $\tilde{\mathcal{L}} \subseteq \mathcal{L}$  to include such nodes.

**Definition 9.** Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ , and let  $\mathcal{R} = \{\mathcal{R}_\ell \mid \ell \in \mathcal{L}\}$  be the set of its confined sensors.

Then  $N$  *selectively propagates confined data* with respect to  $\mathcal{R}$  and  $\tilde{\mathcal{L}}$  if and only if  $N \rightarrow^* N'$  and for all  $\ell_1, \ell_2 \in \mathcal{L}$  there is no transition  $N' \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N''$  such that  $\mathcal{O}(v_{i \downarrow 2}) = \text{confined}$  and  $(\ell_1 \notin \tilde{\mathcal{L}} \text{ or } \ell_2 \notin \tilde{\mathcal{L}})$ , for some  $i \in [1, r]$ .

Once again, the component  $\kappa$  allows us to statically detect when a system of nodes  $N$  selectively propagates confined data within a sub-system, as an instance of Theorem 5.1.

**Theorem 5.6** (Selective data propagation). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ ; let  $\mathcal{R} = \{\mathcal{R}_\ell \mid \ell \in \mathcal{L}\}$  be the set of its confined sensors; and let  $\tilde{\mathcal{L}} \subseteq \mathcal{L}$  be a sub-system. Then  $N$  selectively propagates confined data with respect to  $\mathcal{R}$  and  $\tilde{\mathcal{L}}$  if*

- (1)  $(\hat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then whenever  $\mathcal{O}(\hat{v}_i) = \text{confined}$  for some  $i \in [1, r]$ , it is  $\ell_1, \ell_2 \in \tilde{\mathcal{L}}$ .

Back to our example, let the set of confined sensors  $\mathcal{R}$  be the singleton  $\{S_{cp}\}$ , the set of allowed nodes  $\tilde{\mathcal{L}}$  be  $\{\ell_{cp}, \ell_a, \ell_{pd}\}$  and the set of anonymisation function  $\tilde{\mathcal{F}}$  be  $\{an\}$ . From our analysis we know that  $\kappa(\ell_s) \in (\ell_a, \langle\langle X, Y \rangle\rangle)$ , where  $(\text{CAR}, \{\text{CAR} \rightarrow \text{car}\}) \in X$  and  $\nu \in Y$  where  $\nu$  is defined in (4.2). We discover a policy violation since  $\mathcal{O}(Y) = \text{confined}$  but  $\ell_s \notin \tilde{\mathcal{L}}$ . Now consider our example with the node  $N'_a$ , introduced at the end of Section 2. In this case we have no policy violation because, from our analysis, we obtain that  $\kappa(\ell_s) \in (\ell'_a, \langle\langle X, Z \rangle\rangle)$  where  $X$  is as above, and  $Y$  includes

$$(\text{AN}, \{\text{AN} \rightarrow \text{an}^{\ell'_a}(\text{NOISERED}^{\ell_{cp}}), \text{NOISERED}^{\ell_{cp}} \rightarrow \text{noiseRed}^{\ell_{cp}}(M^{\ell_{cp}}), M^{\ell_{cp}} \rightarrow 1^{\ell_{cp}}\})$$

and  $\mathcal{O}(Y) = \text{open}$ .

Above, we only have two tags, but one can easily adapt the definition taking into account a set of tags. Accordingly, the system of nodes can be partitioned in many sub-systems. It is easy to constrain the propagation of a piece of information with a certain tag to only occur within given sub-nets. A further generalisation refines the data propagation policy by defining for each node  $N_\ell$  and each tag  $d$  the set of nodes to which  $N_\ell$  is allowed to send data tagged with  $d$ .

## 6. CONCLUSIONS

We proposed the process calculus IoT-LySA as a formal design framework for IoT. It aims at supporting the designer of IoT applications and at providing techniques for checking their properties. IoT-LySA has constructs to describe sensors and actuators, and suitable primitives for managing the coordination and the communication capabilities of interconnected smart objects. We equipped our calculus with a Control Flow Analysis that statically computes a safe over-approximation of system behaviour. In particular it predicts how nodes interact, how data spread from sensors to the network, and how data are processed.

These approximations offer the basis for checking various properties of IoT systems, among which the classical security properties of secrecy and no-read-up/no-write-down. In addition, we considered a combination of policies ruling access control and data propagation. In order to better assess the feasibility of our proposal, we are implementing a tool to assist the designer in defining the properties of interest, e.g. following the schema of Section 5, and in checking them on the system at hand. Prior to that is the computation of the (minimal) valid estimate. This requires deriving a set of constraints from the analysis specification and using the succinct solver of [39].

In order to experiment on the usability of IoT-LySA as a specification language we are currently considering a few case studies. In a scenario of precision agriculture, we are modelling an efficient and sustainable irrigation system to manage water in a vineyard [9]. Another example considers an IoT system that controls the temperature in a smart storehouse with particular attention to countermeasures in case an attacker tampers with sensors [15]. We recently started a collaboration with Zerynth, a spin-off of the Pisa University<sup>2</sup> for designing a smart fridge system.

Further future work concerns the extension of IoT-LySA itself. A first issue is considering mobility of smart objects. A simple solution would be making dynamic the topology of the network through suitable primitives that update the compatibility function *Comp* at run time. However, the accuracy of our present analysis would be certainly affected by the arising non-determinism. A richer solution would be representing also spatial information in the code and providing the designer with primitives that specify the movement of objects from one location to another.

Addressing mobility makes it evident that smart objects should adapt their behaviour according to the different features of the location hosting them. Indeed, in the IERC words: “there is still a lack of research on how to adapt and tailor existing research on autonomic computing to the specific characteristics of IoT” [30]. To contribute to these issues, we plan to extend our calculus with linguistic mechanisms and a verification machinery to deal with adaptivity in the style of [22, 12].

As it is, IoT-LySA assumes that the components of a node never fail. To relax this assumption, a quantitative semantics will help in taking into account the probability of failure of each node component. Of course, the analysis and the verification will be deeply affected by this new semantics. Actually, a quantitative semantics would support tuning the design of a system. For example, a designer can better shape the dimension of a node with respect to energy consumption, or balance the computational load of nodes by allocating specific data aggregations within the system. A quantitative semantics would also help in studying the trade-off between the costs of protecting some assets in IoT system and the damage an attacker can cause.

---

<sup>2</sup><https://www.zerynth.com/>

For the time being, IoT-LYSA models only the logical part of a system, leaving the physical world as a black box. A further extension consists in opening this box and in modelling the world through standard techniques as e.g. hybrid automata. This extension not only improves the precision of IoT-LYSA models but also could allow designers to understand how sensors are used and how an action triggered on an actuator affects the whole system. As an example, one can predict if an actuator is maliciously triggered by an attacker, as happened in the recent attack performed through a vehicular infotainment network.<sup>3</sup>

**6.1. Related work.** As all new paradigms, IoT poses new challenging scenarios for formal methodologies. To the best of our knowledge, few papers address the specification and verification of IoT systems from a process calculi perspective.

A first proposal is the IoT-calculus [32]. It explicitly includes sensors and actuators, and smart objects are represented as point-to-point communicating nodes of networks, represented as a graph. Differently from ours, their interconnection topology can vary at run time, when the system interacts with the environment. This is rendered by a first kind of transitions, the others taking care of the node activities. The authors propose two notions of bisimilarity, each based on the different kinds of transitions, that capture system behaviour from the point of view of end-users and of the selected devices. The clean algebraic presentation enables compositional reasoning on systems.

The timed process calculus CIoT [34] specifies physical and logical components, addresses both timing and topology constraints, and allows for node mobility. Furthermore, communications are either short-range or internet-based. The semantics is given on terms of a labelled transition system with transitions that represent the interactions with the physical environment and transitions that represent the node actions. The focus of this paper is mainly on an extensional semantics that emphasises the interaction of IoT systems with the environment, and that provides a *fully abstract* characterisation of the proposed contextual equivalence. The hybrid process calculus CCPS [35] describes both the cyber and the physical aspects of systems (state variables, physical devices, evolution laws, etc.). It has a clearly defined behavioural semantics, based on a transition system and a notion of bisimulation. As in our case, a cyber component governs the interactions with sensors and actuators, and the communications with other cyber components. The physical world is modelled by a tuple of suitable functions over the real numbers that describe how it changes.

Many design choices of the above-discussed proposals are similar to ours. The main difference is that our coordination model is based on a shared store à la Linda instead of a message-based communication à la  $\pi$ -calculus. Furthermore, differently from [32, 34], we are here mainly interested in developing a design framework that includes a static semantics to support various verification techniques and tools for checking properties of IoT applications. In addition, as said above, here we are only interested in modelling the logical components, in determining their boundaries with the physical world, and in abstractly representing the interactions between them. Modelling the physical environment is left as future work, e.g. along the lines suggested in [35]. Finally, we opted for an asynchronous multi-party communication modality among nodes, while in IoT-calculus and in CIoT internode communications are synchronous and point-to-point.

---

<sup>3</sup>See <http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/>

The calculi above and ours are built upon those previously introduced for wireless, sensor and ad hoc networks ([33, 43, 38] to cite only a few). In particular, the calculus in [38] is designed to model so-called broadcast networks, with a dynamically changing topology. It presents some features very similar to ours: an asynchronous local broadcast modality, while intra-node communication relies on a local tuple space. Also, the analysis of the behaviour of broadcast networks is done by resorting to a multi-step static machinery.

Also *aggregate programming* has been proposed for the design and implementation of complex IoT software systems [3], through the *Field calculus* [21], a functional language with constructs to model computation and interaction among large numbers of devices. Aggregate programming overcomes the single-device viewpoint, by adopting a cooperating collection of devices as basic computing unit. Also the rule-based approach of [17] describes a network of sensors and actuators as a distributed collaborative system. It chiefly focusses on globally coordinating the activities of the devices and on controlling the events raised by sensors and the effects of the actions of actuators they trigger. Our approach is complementary to theirs, because we are mainly interested in how the data gathered from sensors are communicated and subsequently aggregated by the nodes in the network. The goals of ThingML [37] are similar to ours in that its authors propose a modelling language for IoT. However, our work is oriented towards formal verification, while theirs is devoted to tools for supporting the development of applications, within a Software Engineering approach.

Security in IoT and cyber-physical systems has been addressed from a process algebraic perspective by a certain number of papers, e.g. [45, 44, 2]. An IoT protocol (the MQ Telemetry Transport 3.1) is formally model in terms of a timed message-passing process algebra [2]. The protocol is statically analysed to understand the robustness of its behaviour in scenarios with different quality of service. The Applied Quality Calculus is used in [44] to study the trade-off between the security requirements and the broadcast communications in wireless-based Cyber-Physical Systems, where nodes have limited capabilities and sensors are poorly reliable. In this proposal, standard cryptographic mechanisms are modelled via term rewriting. Furthermore it is possible to reason about denial of service, because the calculus has explicit notions of failing and unwanted communications.

Very similar to ours is the work in [45], where the key establishment protocol of ZigBee [46] is specified in LYSA [8, 16] and statically analysed with a CFA. They discovered a security flaw arising because freshness is not guaranteed, and proposed a fix.

A variant of the CFA presented here is proposed in [14] for tracking sensitive and untrustworthy data. Taint analysis is used for marking data and for monitoring their propagation at run time across an IoT system so as to check whether those computations considered security critical are not affected by tainted data.

Finally, security is studied in [15] from an “economical” perspective, by providing an enhanced version of IOT-LYSA to infer quantitative measures of cryptographic mechanisms. This paper provides the means for estimating the costs of possible counter-measures to several security risks that may arise in IoT systems. A game theoretical approach to the same problem is in [42]. The authors compute a resource allocation plan as a Pareto-optimal solution, and estimate its efficiency with respect to the energy consumption of the security infrastructure and the costs of its components.

## REFERENCES

- [1] Abadi, M., Gordon, A.: A calculus for cryptographic protocols: The Spi calculus. In: Proc. of the 4th ACM Conference on Computer and Communications Security. pp. 36–47. CCS '97, ACM (1997)
- [2] Aziz, B.: A formal model and analysis of an IoT protocol. *Ad Hoc Networks* 36, 49–57 (2016)
- [3] Beal, J., Pianini, D., Viroli, M.: Aggregate programming for the Internet of Things. *IEEE Computer* 48(9), 22–30 (2015)
- [4] Bell, D., LaPadula, L.: Secure computer systems: Unified exposition and Multics interpretation. Tech. Rep. ESD-TR-75-306, Mitre C. (1976)
- [5] Bertino, E.: Data security and privacy in the IoT. In: Pitoura, E., Maabout, S., Koutrika, G., Marian, A., Tanca, L., Manolescu, I., Stefanidis, K. (eds.) Proceedings of the 19th International Conference on Extending Database Technology. pp. 1–3. OpenProceedings.org (2016)
- [6] Bodei, C., Brodo, L., Focardi, R.: Static evidences for attack reconstruction. In: Programming Languages with Applications to Biology and Security. LNCS, vol. 9465, pp. 162–182. Springer (2015)
- [7] Bodei, C., Buchholtz, M., Degano, P., Nielson, F., Nielson, H.R.: Automatic validation of protocol narration. In: Computer Security Foundations Workshop (CSFW-16 2003). pp. 126–140. IEEE Computer Society (2003)
- [8] Bodei, C., Buchholtz, M., Degano, P., Nielson, F., Nielson, H.R.: Static validation of security protocols. *Journal of Computer Security* 13(3), 347–390 (2005)
- [9] Bodei, C., Degano, P., Ferrari, G.L., Galletta, L.: Sustainable precision agriculture from a process algebraic perspective: A smart vineyard, submitted for publication, preprint available at <https://gofile.io/?c=hbV2W3>
- [10] Bodei, C., Degano, P., Ferrari, G.L., Galletta, L.: A step towards checking security in IoT. In: Proc. of ICE 2016. EPTCS, vol. 223, pp. 128–142 (2016)
- [11] Bodei, C., Degano, P., Ferrari, G.L., Galletta, L.: Where do your IoT ingredients come from? In: Proc. of Coordination 2016. LNCS, vol. 9686, pp. 35–50. Springer (2016)
- [12] Bodei, C., Degano, P., Galletta, L., Salvatori, F.: Context-aware security: Linguistic mechanisms and static analysis. *Journal of Computer Security* 24 (4), 427–477 (2016)
- [13] Bodei, C., Degano, P., Nielson, F., Nielson, H.R.: Static analysis for the  $\pi$ -calculus with applications to security. *Information and Computation* 168, 68–92 (2001)
- [14] Bodei, C., Galletta, L.: The cost of securing IoT communications? In: Proc. of Italian Conference on Cybersecurity (ITASEC17). CEUR Proceedings Vol-1816 (2016)
- [15] Bodei, C., Galletta, L.: Tracking Sensitive and Untrustworthy Data in IoT. In: Proc. of Italian Conference on Theoretical Computer Science. CEUR Proceedings Vol-1720 (2016)
- [16] Buchholtz, M., Nielson, H.R., F. Nielson, F.: A calculus for control flow analysis of security protocols. *International Journal of Information Security* 2(3), 145–167 (2004)
- [17] Cano, J., Rutten, É., Delaval, G., Benazzouz, Y., Gürgen, L.: ECA rules for iot environment: A case study in safe design. In: Eighth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshops, SASOW 2014, London, United Kingdom, September 8-12, 2014. pp. 116–121 (2014)
- [18] Carriero, N., Gelernter, D.: A computational model of everything. *Commun. ACM* 44(11), 77–81 (2001)
- [19] Comon, H., Dauchet, M., Gilleron, R., Löding, C., Jacquemard, F., Lugiez, D., Tison, S., Tommasi, M.: Tree automata techniques and applications. Available on: <http://www.grappa.univ-lille3.fr/tata> (2007), release October, 12th 2007
- [20] Cousot, P., Cousot, R.: Abstract Interpretation: A Unified Lattice Model for Static Analysis of Programs by Construction or Approximation of Fixpoints. In: Proceedings of the 4th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages. pp. 238–252. ACM, New York, NY, USA (1977)
- [21] Damiani, F., Viroli, M., Pianini, D., Beal, J.: Code mobility meets self-organisation: A higher-order calculus of computational fields. In: Proc. of Formal Techniques for Distributed Objects, Components, and Systems - 35th IFIP WG 6.1 International Conference, FORTE 2015. Lecture Notes in Computer Science, vol. 9039, pp. 113–128. Springer (2015)
- [22] Degano, P., Ferrari, G.L., Galletta, L.: A Two-Component Language for Adaptation: Design, Semantics, and Program Analysis. *IEEE Trans. Software Eng.* 42(6), 505 – 529 (2016)
- [23] Denning, D.: *Cryptography and Data Security*. Addison-Wesley, Reading Mass. (1982)
- [24] Elejoste, P., Perallos, A., Chertudi, A., Angulo, I., Moreno, A., Azpilicueta, L., Astrain, J., Falcone, F., Villadangos, J.E.: An easy to deploy street light control system based on wireless communication and led technology. *Sensors* 13(5), 6492–6523 (2013)

- [25] Escolar, S., Carretero, J., Marinescu, M., Chessa, S.: Estimating energy savings in smart street lighting by using an adaptive control system. *IJDSN 2014* (2014)
- [26] Gelernter, D.: Generative communication in Linda. *ACM Trans. Program. Lang. Syst.* 7(1), 80–112 (1985)
- [27] Heintze, N.: Set-based analysis of ML programs. In: *Proceedings of the 1994 ACM conference on LISP and functional programming*. pp. 306–317. LFP '94, ACM, New York, NY, USA (1994)
- [28] Henzinger, T.A.: The theory of hybrid automata. In: *Proceedings, 11th Annual IEEE Symposium on Logic in Computer Science*. pp. 278–292. IEEE Computer Society (1996)
- [29] Herlihy, M.: Wait-free synchronization. *ACM Trans. Program. Lang. Syst.* 13(1), 124–149 (1991)
- [30] IERC: The Internet of Things 2012 – New Horizons (2012), [http://www.internet-of-things-research.eu/pdf/IERC\\_Cluster\\_Book\\_2012\\_WEB.pdf](http://www.internet-of-things-research.eu/pdf/IERC_Cluster_Book_2012_WEB.pdf)
- [31] Khedker, U., Sanyal, A., Karkare, B.: *Data Flow Analysis: Theory and Practice*. CRC Press, Inc., Boca Raton, FL, USA, 1st edn. (2009)
- [32] Lanese, I., Bedogni, L., Felice, M.D.: Internet of things: a process calculus approach. In: *Proc of the 28th Annual ACM Symposium on Applied Computing, SAC '13*. pp. 1339–1346. ACM (2013)
- [33] Lanese, I., Sangiorgi, D.: An operational semantics for a calculus for wireless systems. *Theor. Comput. Sci.* 411(19), 1928–1948 (2010)
- [34] Lanotte, R., Merro, M.: A semantic theory of the internet of things - (extended abstract). In: *Procs. of Coordination 2016*. LNCS, vol. 9686, pp. 157–174. Springer (2016)
- [35] Lanotte, R., Merro, M.: A calculus of cyber-physical systems. In: *Proc. of Language and Automata Theory and Applications (LATA 2017)*. *Lecture Notes in Computer Science*, vol. 10168, pp. 115–127 (2017)
- [36] Milner, R., Parrow, J., Walker, D.: A calculus of mobile processes, I. *Inf. Comput.* 100(1), 1–40 (1992)
- [37] Morin, B., Harrand, N., Fleurey, F.: Model-based software engineering to tame the IoT jungle. *IEEE Software* 34(1), 30–36 (2017)
- [38] Nanz, S., Nielson, F., Nielson, H.R.: Static analysis of topology-dependent broadcast networks. *Inf. Comput.* 208(2), 117–139 (2010)
- [39] Nielson, F., Nielson, H.R., Seidl, H.: A succinct solver for ALFP. *Nordic J. of Computing* 9(4), 335–372 (Dec 2002)
- [40] Nielson, H.R., Nielson, F.: Flow logic: A multi-paradigmatic approach to static analysis. In: *The Essence of Computation, Complexity, Analysis, Transformation*. LNCS, vol. 2566, pp. 223–244. Springer (2002)
- [41] Pierce, B.C.: *Types and programming languages*. MIT Press, Cambridge, MA, USA (2002)
- [42] Rullo, A., Midi, D., Serra, E., Bertino, E.: Strategic security resource allocation for internet of things. In: *36th IEEE International Conference on Distributed Computing Systems*. pp. 737–738. IEEE Computer Society (2016)
- [43] Singh, A., Ramakrishnan, C.R., Smolka, S.: A process calculus for mobile ad hoc networks. *Sci. Comput. Program.* 75(6), 440–469 (2010)
- [44] Vigo, R., Nielson, F., Nielson, H.R.: Broadcast, denial-of-service, and secure communication. In: *Proc. of Integrated Formal Methods, 10th International Conference, IFM 2013*. *Lecture Notes in Computer Science*, vol. 7940, pp. 412–427. Springer (2013)
- [45] Yüksel, E., Nielson, H.R., Nielson, F.: A secure key establishment protocol for ZigBee wireless sensor networks. *Comput. J.* 54(4), 589–601 (2011)
- [46] ZigBee Alliance: <http://www.zigbee.org/>



## APPENDIX A. NOTATION

$N \in \mathcal{N}$	system of nodes $N$	$v \in \mathcal{V}$	values
$\ell \in \mathcal{L}$	the set of labels	$x \in \mathcal{X}_\ell$	variables of node $\ell$
$\ell : [B]$	node at $\ell$ with component $B$	$f \in \mathcal{F}$	functions on terms $E$
$S, A$	sensors and actuators	$\llbracket E \rrbracket_\Sigma$	denotational semantics of terms
$i \in \mathcal{I}_\ell, j \in \mathcal{J}_\ell$	unique identifiers of $S$ and $A$	$\mathcal{E}$	state of the world
$\Sigma_\ell : \mathcal{X}_\ell \cup \mathcal{I}_\ell \rightarrow \mathcal{V}$	store of node $\ell$	$\xrightarrow{\mathcal{E}}$	reduction relation
$\widehat{G} = (\mathbb{N}, \mathbb{T}, Z, R)$	regular tree grammar	$\xrightarrow{\mathcal{E}'}$	abstract sensor, value, function
$Lang(\widehat{G})$	language of $\widehat{G}$	$i^\ell, v^\ell, f^\ell \in \mathbb{T}$	tree over $\mathbb{T}$
$\hat{v} = (Z, R)$	shorthand for $\widehat{G}$	$\hat{t} \in \mathcal{T}$	abstract terms
$\mathbb{R}$	the productions over $\mathbb{N}, \mathbb{T}$	$\widehat{\mathcal{V}} = 2^{\mathbb{N} \times \mathbb{R}}$	abstract store at $\ell$
$(\widehat{\Sigma}, \Theta)$	estimate for terms $E$	$\widehat{\Sigma}_\ell : \mathcal{X} \cup \mathcal{I}_\ell \rightarrow 2^{\widehat{\mathcal{V}}}$	abstract store
$(\widehat{\Sigma}, \kappa, \Theta)$	estimate for nodes $N$	$\widehat{\Sigma} = \bigcup_{\ell \in \mathcal{L}} \widehat{\Sigma}_\ell$	abstract network environment
$\vartheta \in 2^{\widehat{\mathcal{V}}}$	set of approximated values	$\kappa : \mathcal{L} \rightarrow \mathcal{L} \times \bigcup_{i=1}^m \widehat{\mathcal{V}}^i$	abstract data collection
$Comp(\ell_2, \ell_1)$	compatibility function	$\Theta : \mathcal{L} \rightarrow 2^{\widehat{\mathcal{V}}}$	instrumented store
$\llbracket E \rrbracket_{\Sigma_\ell}^i$	instrumented semantics	$\Sigma_\ell^i : \mathcal{X}_\ell \cup \mathcal{I}_\ell \rightarrow \mathcal{V} \times \widehat{\mathcal{T}}$	concrete and abstract store agree
$k \in \mathcal{K}$	crypto key	$\Sigma_\ell^i \bowtie \widehat{\Sigma}_\ell$	abstract values for security
$\mathfrak{D}$	set of security tags	$enc_r^\ell, k_i$	tagging functions
$\mathfrak{P} \subseteq \mathfrak{D} \times \mathcal{L} \times \mathcal{L}$	security policy	$\mathfrak{T}_\mathfrak{D} : \widehat{\mathcal{T}} \rightarrow \mathfrak{D}$ and	
		$\mathfrak{T}_\mathfrak{S} : \widehat{\mathcal{V}} \rightarrow \mathfrak{D}$	

Table 4: Notation and abbreviations

## APPENDIX B. PROOFS

In this appendix, we restate the theorems presented earlier in the paper and give the proofs of their correctness, with the help of some auxiliary lemmata.

Our first lemma guarantees that the analysis for expressions respects the instrumented denotational semantics.

**Lemma 1** (Subject reduction for expressions).

For all  $E$ ,  $(\widehat{\Sigma}, \Theta) \models_\ell E : \vartheta$  implies that there exist  $\Sigma_\ell^i$  and  $\widehat{G} \in \vartheta$  such that  $\Sigma_\ell^i \bowtie \widehat{\Sigma}_\ell$  and  $(\llbracket E \rrbracket_{\Sigma_\ell^i}^i)_{\downarrow 2} \in Lang(\widehat{G})$ .

*Proof.* A straightforward induction on all the rules in Table 2 and on rule (E-enc) suffices.  $\square$

Before proving the correctness of the analysis for systems of nodes, we prove that it is invariant under the structural congruence and is preserved under node components reduction.

**Lemma 2** (Congruence).

- If  $B \equiv B'$  then  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B$  iff  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B'$ .
- If  $N \equiv N'$  then  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  iff  $(\widehat{\Sigma}, \kappa, \Theta) \models N'$ .

*Proof.* It suffices to inspect the rules for  $\equiv$ , since associativity and commutativity of  $\wedge$  reflects the same properties of both  $|$  and  $\|$ , and to recall that any triple is a valid estimate for  $0$ .  $\square$

**Lemma 3** (Subject reduction for node components).

If  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} B$  and  $B \xrightarrow[\mathcal{E}']{\mathcal{E}} B'$  and  $\Sigma_{\ell}^i \bowtie \widehat{\Sigma}_{\ell}$ , for  $\Sigma_{\ell}^i$  (in  $B$ ) then  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} B'$  and  $\Sigma_{\ell}^{i'} \bowtie \widehat{\Sigma}_{\ell}$ , for  $\Sigma_{\ell}^{i'}$  (in  $B'$ ).

*Proof.* Our proof is by induction on the shape of the derivation of  $B \xrightarrow[\mathcal{E}']{\mathcal{E}} B'$  and by cases on the last rule used. In all the cases below we will have that for  $\Sigma^i$  and  $B_1$  components of  $B$ , both facts (\*)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i$ , and (\*\*)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} B_1$  trivially hold. So we will omit mentioning these judgements below, for the sake of brevity. Also, we will implicitly apply the rules of Table 3, e.g. the repeated applications of rule (B-par).

- Case (Sense). We assume  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \parallel \text{probe}(i).S \parallel B_1$  that has been proved because rule (B-sen) proves that  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \text{probe}(i).S$  holds. By the same rule also  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} S$  holds, and thus  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \{(v, i^{\ell})/i\} \parallel S \parallel B_1$ . We are left to show that  $\Sigma_{\ell}^{i'} \bowtie \widehat{\Sigma}_{\ell}$ . By hypothesis  $\Sigma_{\ell}^{i'}(y) = \Sigma_{\ell}^i(y)$  for all  $y \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  such that  $y \neq i$ , while  $(\Sigma_{\ell}^i(i))_{\downarrow 2} = i^{\ell} \in \text{Lang}(\widehat{G})$ , for some  $\widehat{G} \in \widehat{\Sigma}_{\ell}(i)$ , which trivially holds by (\*) and Lemma 1.
- Case (Asgm). We assume  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \parallel x := E.P \parallel B_1$  that has been proved because the following conditions hold:

$$(\widehat{\Sigma}, \Theta) \models_{\ell} E : \vartheta \tag{B.1}$$

$$\forall \hat{v} : \hat{v} \in \vartheta \Rightarrow \hat{v} \in \widehat{\Sigma}_{\ell}(x) \tag{B.2}$$

$$(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} P \tag{B.3}$$

We have to prove that  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \{(v, \hat{v})/x\} \parallel P \parallel B_1$  that trivially holds because of (\*); (B.3) and Lemma 1; and (\*\*). We have only to prove that  $\Sigma_{\ell}^{i'} \bowtie \widehat{\Sigma}_{\ell}$ , by knowing that (by hypothesis)  $\Sigma_{\ell}^i \bowtie \widehat{\Sigma}_{\ell}$ , that implies  $\Sigma_{\ell}^{i'}(y) = \Sigma_{\ell}^i(y)$  for all  $y \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  such that  $y \neq x$ . Therefore  $\Sigma_{\ell}^{i'}(y) = \Sigma_{\ell}^i(y)$  for all  $y \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  such that  $y \neq x$ , while  $(\Sigma_{\ell}^i(x))_{\downarrow 2} \in \text{Lang}(\hat{v} \in \widehat{\Sigma}_{\ell}(x))$  because of (B.2) and of Lemma 1.

- Case (E-out). We assume  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \parallel \langle\langle E_1, \dots, E_r \rangle\rangle \triangleright L.P \parallel B_1$  that has been proved because the following conditions hold:

$$\begin{aligned} & \bigwedge_{i=1}^r (\widehat{\Sigma}, \Theta) \models_{\ell} E_i : \vartheta_i \\ & (\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} P \\ & \forall \hat{v}_1, \dots, \hat{v}_r : \bigwedge_{i=1}^r \hat{v}_i \in \vartheta_i \Rightarrow \forall \ell' \in L : (\ell, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell') \end{aligned}$$

Proving that  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L.0 \parallel P \parallel B_1$  is straightforward because of the three above conditions and because of Lemma 1.

- Case (Decrypt). We assume  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \parallel \text{decrypt } E \text{ as } \{E'_1, \dots, E'_j; x_{j+1}, \dots, x_r\}_k \text{ in } P \parallel B_1$  that has been proved because the following conditions hold

$$(\widehat{\Sigma}, \Theta) \models_{\ell} E : \vartheta \tag{B.4}$$

$$\bigwedge_{i=1}^r (\widehat{\Sigma}, \Theta) \models_{\ell} E_i : \vartheta_i \quad (\text{B.5})$$

and  $\forall \hat{v} \in \vartheta$  such that  $D(\hat{v}, k) = [\hat{v}_1, \dots, \hat{v}_r]$  that also implies the following two conditions

$$\bigwedge_{i=j+1}^r \hat{v}_i \in \widehat{\Sigma}_{\ell}(x_i) \quad (\text{B.6})$$

$$(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} P \quad (\text{B.7})$$

We have to prove that  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i \parallel P[v_{j+1}/x_{j+1}, \dots, v_r/x_r] \parallel B_1$  that trivially holds because of (\*); (B.4), (B.5), (B.7) and Lemma 1; and (\*\*). We are left to prove that  $\Sigma_{\ell}^{i'} \bowtie \widehat{\Sigma}_{\ell}$ , by knowing that (by hypothesis)  $\Sigma_{\ell}^i \bowtie \widehat{\Sigma}_{\ell}$ , that implies  $\Sigma_{\ell}^{i'}(y) = \Sigma_{\ell}^i(y)$  for all  $y \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  such that  $y \neq x$ . Therefore  $\Sigma_{\ell}^{i'}(y) = \Sigma_{\ell}^i(y)$  for all  $y \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell}$  such that  $y \neq x$ , while  $(\Sigma_{\ell}^i(x))_{\downarrow 2} \in \text{Lang}(\hat{v} \in \widehat{\Sigma}_{\ell}(x))$  because of (B.4), (B.5) and because of Lemma 1.

- The cases (Cond), (Int) and (A-com) are straightforward; the case (Act) is trivial because for us the world is a black box; the case (ParB) directly follows from the induction hypothesis; Lemma 2 suffices to prove the case (CongrB).  $\square$

**Theorem 4.1** (Subject reduction). *Given a system of nodes  $N$ , if  $(\widehat{\Sigma}, \kappa, \Theta) \models N; N \xrightarrow[\mathcal{E}']{\mathcal{E}} N'$  and  $\forall \Sigma_{\ell}^i$  in  $N$  it is  $\Sigma_{\ell}^i \bowtie \widehat{\Sigma}_{\ell}$ , then  $(\widehat{\Sigma}, \kappa, \Theta) \models N'$  and  $\forall \Sigma_{\ell}^{i'}$  in  $N'$  it is  $\Sigma_{\ell}^{i'} \bowtie \widehat{\Sigma}_{\ell}$ .*

*Proof.* Our proof is by induction on the shape of the derivation of  $N \xrightarrow[\mathcal{E}']{\mathcal{E}} N'$  and by cases on the last rule used. In all the cases below we will have that (\*)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} \Sigma^i$ , as well as that (\*\*)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell} B_m$  (whenever  $m \geq 1$ ), for  $\Sigma^i, B_m$  components of  $N$ . So we will omit mentioning these judgements below, for the sake of brevity. Also, we will implicitly apply the rules of Table 3, e.g. the applications of rules (N-node) and (N-par).

- Case (Multi-com). We assume

$$(\widehat{\Sigma}, \kappa, \Theta) \models \ell_1 : [\langle v_1, \dots, v_r \rangle \triangleright L.0 \parallel B_1] \mid \ell_2 : [\Sigma_{\ell_2}^i \parallel (E_1, \dots, E_j; x_{j+1}, \dots, x_r).Q \parallel B_2]$$

that is implied by  $(\widehat{\Sigma}, \kappa, \Theta) \models \ell_1 : [\langle v_1, \dots, v_r \rangle \triangleright L.0 \parallel B_1]$  and by

$(\widehat{\Sigma}, \kappa, \Theta) \models \ell_2 : [\Sigma_{\ell_2}^i \parallel (E'_1, \dots, E'_j; x_{j+1}, \dots, x_r).Q \parallel B_2]$  that have been proved because the following conditions hold:

$$\bigwedge_{i=1}^r (\widehat{\Sigma}, \Theta) \models_{\ell_1} v_i : \vartheta_i \quad (\text{B.8})$$

$$(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_1} 0 \quad (\text{B.9})$$

$$\forall \hat{v}_1, \dots, \hat{v}_r : \bigwedge_{i=1}^r \hat{v}_i \in \vartheta_i \Rightarrow \forall \ell' \in L : (\ell_1, \langle \hat{v}_1, \dots, \hat{v}_r \rangle) \in \kappa(\ell') \quad (\text{B.10})$$

$$\bigwedge_{i=1}^j (\widehat{\Sigma}, \Theta) \models_{\ell_2} E_i : \vartheta'_i \quad (\text{B.11})$$

$$(\text{B.12})$$

and  $\forall (\ell_1, \langle \hat{v}_1, \dots, \hat{v}_k \rangle) \in \kappa(\ell_2)$  such that  $\text{Comp}(\ell_1, \ell_2)$  ( $\clubsuit$ )

$$\bigwedge_{i=j+1}^r \hat{v}_i \in \widehat{\Sigma}_{\ell_2}(x_i) \quad (\text{B.13})$$

$$(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_2} Q \quad (\text{B.14})$$

Note that since  $\clubsuit$  holds, also B.13 and B.14 do. Also,  $\forall i (\widehat{\Sigma}, \Theta) \models_{\ell_1} v_i : \vartheta_i$  implies  $\hat{v}_i \in \vartheta_i$ , where  $\hat{v}_i = (\llbracket v_i \rrbracket_{\Sigma_{\ell_2}^i}^i)_{\downarrow 2}$ , and that  $\ell_2 \in L$  because  $N \xrightarrow[\mathcal{E}']{\mathcal{E}} N'$ . We have to prove that

$$(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_1} : [\langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L \setminus \{\ell_2\}.0 \parallel B_1] \mid \ell_2 : [\Sigma_{\ell_2}^i \{v_{j+1}/x_{j+1}, \dots, v_r/x_r\} \parallel Q \parallel B_2]$$

that in turn amounts to prove that

- (a)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_1} \langle\langle v_1, \dots, v_r \rangle\rangle \triangleright L \setminus \{\ell_2\}.0 \parallel B_1$
- (b)  $(\widehat{\Sigma}, \kappa, \Theta) \models_{\ell_2} \Sigma_{\ell_2}^i \{(v_{j+1}, \hat{v}_{j+1})/x_{j+1}, \dots, (v_r, \hat{v}_r)/x_r\} \parallel Q \parallel B_2$

We have that (a) holds trivially because of (B.8), (B.9) and (B.10) (of course  $L \setminus \{\ell_2\} \subseteq L$ ), while (b) holds because of the remaining conditions and of Lemma 1. We are left to prove that  $\Sigma_{\ell_2}^{i'}$   $\bowtie$   $\widehat{\Sigma}_{\ell_2}$ . Now, we know that  $\Sigma_{\ell_2}^{i'}(y) = \Sigma_{\ell_2}^i(y)$  for all  $y \in \mathcal{X}_{\ell_2} \cup \mathcal{I}_{\ell_2}$  such that  $y \neq x_i$ . The condition  $(\Sigma_{\ell_2}^i(x_i))_{\downarrow 2} \in \text{Lang}(\hat{v} \in \widehat{\Sigma}_{\ell_2}(x_i))$  holds for all  $x_i$  because of (B.13).

- The case (CongrN) follows from Lemma 2; and the remaining cases directly follow from the induction hypothesis.  $\square$

**Definition 10.** The set of estimates can be partially ordered by setting

$$(\widehat{\Sigma}_1, \kappa_1, \Theta_1) \sqsubseteq (\widehat{\Sigma}_2, \kappa_2, \Theta_2) \text{ iff}$$

- $\forall x \in \mathcal{X}_{\ell} \cup \mathcal{I}_{\ell} : \widehat{\Sigma}_1(x) \subseteq \widehat{\Sigma}_2(x)$
- $\forall \ell \in \mathcal{L} : \kappa_1(\ell) \subseteq \kappa_2(\ell)$
- $\forall \ell \in \mathcal{L} : \Theta_1(\ell) \subseteq \Theta_2(\ell)$

This suffices for making the set of proposed solutions into a complete lattice; we can thus write  $(\widehat{\Sigma}_1, \kappa_1, \Theta_1) \sqcup (\widehat{\Sigma}_2, \kappa_2, \Theta_2)$  for the binary least upper bound (defined point-wise),  $\sqcap \mathcal{M}$  for the greatest lower bound of a set  $\mathcal{M}$  of proposed solutions (also defined pointwise), and  $(\perp, \perp, \perp)$  for the least element.

A Moore family  $\mathcal{M}$  contains a greatest element  $(\sqcap \emptyset)$  and a least element  $(\sqcap \mathcal{M})$ . Since the set of analysis estimates constitutes a Moore family, we have that there always is a least estimate to the specification in Table 3.

**Theorem 4.2** (Existence of estimates). *Given  $N$ , its valid estimates form a Moore family.*

*Proof.* Let  $\mathcal{M} = \{(\widehat{\Sigma}_r, \kappa_r, \Theta_r)\}$  be a set of estimates for  $N$ . We proceed by structural induction on  $N$ , to check that  $\sqcap \mathcal{M} = (\widehat{\Sigma}', \kappa', \Theta') \models N$ . We just consider one case. The others are similar.

Case  $\ell : [B]$ . Since  $\forall r : (\widehat{\Sigma}_r, \kappa_r, \Theta_r) \models \ell : [B]$ , then  $(\widehat{\Sigma}_r, \kappa_r, \Theta_r) \models_{\ell} B$ .

Using the induction hypothesis and the fact that the components of  $(\widehat{\Sigma}', \kappa', \Theta')$  are obtained pointwise, it follows that  $(\widehat{\Sigma}', \kappa', \Theta') \models_{\ell} B$  thus establishing the required judgement  $(\widehat{\Sigma}', \kappa', \Theta') \models \ell : [B]$ .  $\square$

The following auxiliary lemma helps proving Corollary 4.3, and in its statement we let  $B \xrightarrow{E_1, \dots, E_n}_\ell B'$  denote a reduction in which all  $E_i$  are evaluated at node  $\ell$ , omitting the environment for readability.

**Lemma 4.** *If  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B$  and  $B \xrightarrow{E_1, \dots, E_n}_\ell B'$  then  $\forall r \in [0, n]$  there exists  $\widehat{G} \in \Theta(\ell)$  such that  $(\llbracket E_r \rrbracket_{\Sigma_i}^i)_{\downarrow 2} \in \text{Lang}(\widehat{G})$ .*

*Proof.* By Lemma 3,  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B'$  holds. The proof proceeds by induction on the shape of the derivation of  $B \xrightarrow[\mathcal{E}']{\mathcal{E}} B'$  and by cases on the last rule used.

- Case (Asgm). If this rule is applied, then  $B = \Sigma^i \parallel x := E. P \parallel B_1$ . Since  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B$  we have that, in particular,  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell x := E. P$  and in turn that  $(\widehat{\Sigma}, \Theta) \models_\ell E : \vartheta$ , and Lemma 1 suffices for establishing that  $(\widehat{\Sigma}, \Theta) \models_\ell \llbracket E \rrbracket^i : \vartheta$ . By the rules in Table 2, we have the required  $\vartheta \subseteq \Theta(\ell)$ .
- Case (Decrypt). In this case we have that  $(\widehat{\Sigma}, \kappa, \Theta) \models_\ell B = \Sigma^i \parallel \text{decrypt } E \text{ as } \{E'_1, \dots, E'_j; x_{j+1}, \dots, x_r\}_k \text{ in } P \parallel B_1$ . Since the estimate is valid, we have in particular that the premises of the rule (i.e. conditions (B.4), (B.5), and (B.6) stated in the proof of Lemma 3) hold. The existence of the required  $\widehat{G}$  now follows, because Lemma 1 can be used to recover the abstract values  $\hat{v}, \hat{v}_1, \dots, \hat{v}_n$  needed in the rule (P-dec).
- The cases (Ev-out), and (Cond) are similar.
- The cases (ParB) and (CongrB) directly follow from the induction hypothesis.
- No other rules evaluate terms and their proof is trivial. □

**Corollary 4.3.** *Assume that  $(\widehat{\Sigma}, \kappa, \Theta) \models N$ , then*

- (1) *if  $N \xrightarrow{E_1, \dots, E_r}_\ell N'$  then  $\forall k \in [0, n]$  there exists  $\widehat{G} \in \Theta(\ell)$  such that  $(\llbracket E \rrbracket_{\Sigma_i}^i)_{\downarrow 2} \in \text{Lang}(\widehat{G})$ ;*
- (2) *if  $N \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N'$  then it holds  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$ , and  $\forall i \in [0, r]$  there exists  $\widehat{G} \in \Theta(\ell_2)$  such that  $v_{i\downarrow 2} \in \text{Lang}(\widehat{G})$ .*

*Proof.*

- (1) Immediate; note that for the rule (Node) Lemma 4 suffices.
- (2) By Theorem 4.1, we have that  $(\widehat{\Sigma}, \kappa, \Theta) \models N'$ , so we proceed by induction on the shape of the derivation of  $N \xrightarrow[\mathcal{E}']{\mathcal{E}} N'$  and by cases on the last rule used.
  - Case (Multi-com) follows directly from the subject reduction Theorem 4.1.
  - Cases (ParN), (CongrN), and (Node) directly follow from the induction hypothesis, and for the other rules the premise is false. □

**Proposition 4.4.** *Let  $i$  be a sensor of the node  $N$  with label  $\ell$ ; let  $N'$  be a node with label  $\ell'$ ; and let  $(\widehat{\Sigma}, \kappa, \Theta) \models N \mid N'$ .*

*The sensor  $i$  is an ingredient of  $N'$  if there exists  $\widehat{G} \in \Theta(\ell')$  and a tree in its language with a leaf  $i^\ell$ .*

*Proof.* Immediate. □

**Theorem 5.1** (Well propagation). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ ; let  $\mathcal{D}$  be a finite set of tags; let  $\mathfrak{T}_{\mathcal{D}}, \mathfrak{T}_{\mathcal{G}}$  be a pair of tagging functions; and let  $\mathfrak{P} \subseteq \mathcal{D} \times \mathcal{L} \times \mathcal{L}$ .*

*Then  $N$  enjoys  $\mathfrak{P}$  if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then  $\forall i \in [1, r]$  it holds  $\mathfrak{P}(\mathfrak{T}_{\mathcal{G}}(\hat{v}_i), \ell_1, \ell_2)$ .

*Proof.* Suppose that both items (1) and (2) hold and assume by contradiction that  $N$  enjoys  $\mathfrak{P}$  with respect to  $\mathcal{D}$ , i.e. that  $N \rightarrow^* N' \xrightarrow{\langle\langle v_1, \dots, v_r \rangle\rangle}_{\ell_1, \ell_2} N''$  and that  $\exists \bar{i} : \mathfrak{P}(\mathfrak{T}_{\mathcal{D}}(v_{\bar{i}}), \ell_1, \ell_2)$  does not hold. Then, by Corollary 4.3, it holds  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$ , where for all  $i$  there exists  $\hat{G} \in \hat{v}_i$  such that  $v_{i \downarrow 2} \in \text{Lang}(\hat{G})$ . Now, since item (2) holds, we know that  $\forall i. \mathfrak{P}(\mathfrak{T}_{\mathcal{G}}(\hat{v}_i), \ell_1, \ell_2)$  holds. Since  $\mathfrak{T}_{\mathcal{D}}, \mathfrak{T}_{\mathcal{G}}$  is a pair of tagging functions, for all  $\hat{t} \in \text{Lang}(\hat{G})$  we have that  $\mathfrak{P}(\mathfrak{T}_{\mathcal{G}}(\hat{v}_i), \ell_1, \ell_2) = \mathfrak{P}(\mathfrak{T}_{\mathcal{D}}(\hat{t}), \ell_1, \ell_2)$ , in particular for  $i = \bar{i}$ : contradiction.  $\square$

**Fact 5.2.**  $\mathcal{D}$  and  $\mathcal{S}$  are a pair of tagging functions.

*Proof.* Immediate by inducing on the structure of  $\hat{t}$ .  $\square$

**Theorem 5.3** (Confidentiality). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ , and let  $\mathcal{S} = \{\mathcal{S}_{\ell} \mid \ell \in \mathcal{L}\}$  be the set of its secret sensors. Then  $N$  has no leaks with respect to  $\mathcal{S}$  if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then  $\forall i \in [1, r]$   $\mathcal{S}(\hat{v}_i) = \text{public}$ .

*Proof.* Follows easily from Theorem 5.1 by letting  $\mathcal{D} = \{\text{secret}, \text{public}\}$ ;  $\mathcal{S}, \mathcal{P}$  as pair of tagging functions (Lemma 5.2); and  $\mathfrak{P}(d, \ell_1, \ell_2) = \text{true}$  if and only if  $d = \text{public}$ .  $\square$

**Theorem 5.4** (No-read-up/no-write-down). *Given an assignment function  $\text{level}$ , and a system of nodes  $N$  with labels in  $\mathcal{L}$ . Then  $N$  respects the levels if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then  $\text{level}(\ell_1) \leq \text{level}(\ell_2)$ .

*Proof.* Follows easily from Theorem 5.1 by letting  $\mathcal{D} = \{\bullet\}$ ;  $\mathfrak{T}_{\mathcal{G}}, \mathfrak{T}_{\mathcal{D}}$  as pair of (constant) tagging functions; and  $\mathfrak{P}(d, \ell_1, \ell_2) = \text{true}$  if and only if  $\text{level}(\ell_1) \leq \text{level}(\ell_2)$ .  $\square$

**Fact 5.5.**  $\mathcal{C}$  and  $\mathcal{O}$  are a pair of tagging functions.

*Proof.* Immediate by inducing on the structure of  $\hat{t}$ .  $\square$

**Theorem 5.6** (Selective data propagation). *Let  $N$  be a system of nodes with labels in  $\mathcal{L}$ ; let  $\mathcal{R} = \{\mathcal{R}_{\ell} \mid \ell \in \mathcal{L}\}$  be the set of its confined sensors; and let  $\tilde{\mathcal{L}} \subseteq \mathcal{L}$  be a sub-system. Then  $N$  selectively propagates confined data with respect to  $\mathcal{R}$  and  $\tilde{\mathcal{L}}$  if*

- (1)  $(\widehat{\Sigma}, \kappa, \Theta) \models N$  and
- (2)  $\forall \ell_1, \ell_2 \in \mathcal{L}$  if  $(\ell_1, \langle\langle \hat{v}_1, \dots, \hat{v}_r \rangle\rangle) \in \kappa(\ell_2)$  then whenever  $\mathcal{O}(\hat{v}_i) = \text{confined}$  for some  $i \in [1, r]$ , it is  $\ell_1, \ell_2 \in \tilde{\mathcal{L}}$ .

*Proof.* Follows easily from Theorem 5.1 by letting  $\mathcal{D} = \{\text{confined}, \text{open}\}$ ;  $\mathfrak{T}_{\mathcal{G}} = \mathcal{O}$  and  $\mathfrak{T}_{\mathcal{D}} = \mathcal{C}$  that form a pair of tagging functions by Lemma 5.5; and  $\mathfrak{P}(d, \ell_1, \ell_2) = \text{true}$  if and only if  $d = \text{confined}$  and  $\ell_1, \ell_2 \in \tilde{\mathcal{L}}$ .  $\square$